

Billings, Stephen; Johnson, Erik

Conference Paper

A nonparametric test for industrial specialization

52nd Congress of the European Regional Science Association: "Regions in Motion - Breaking the Path", 21-25 August 2012, Bratislava, Slovakia

Provided in Cooperation with:

European Regional Science Association (ERSA)

Suggested Citation: Billings, Stephen; Johnson, Erik (2012) : A nonparametric test for industrial specialization, 52nd Congress of the European Regional Science Association: "Regions in Motion - Breaking the Path", 21-25 August 2012, Bratislava, Slovakia, European Regional Science Association (ERSA), Louvain-la-Neuve

This Version is available at:

<https://hdl.handle.net/10419/120611>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.

A Nonparametric Test for Industrial Specialization

Stephen B. Billings*
and Erik B. Johnson†

February 16, 2012

Abstract

We introduce a nonparametric microdata based test for industrial specialization and apply it to a single urban area. Our test employs establishment densities for specific industries, a population counterfactual, and a new correction for multiple hypothesis testing to determine the statistical significance of specialization across both places and industries. Results highlight patterns of specialization that are extremely varied, with downtown places specializing in a number of service sector industries, while suburban places specialize in both manufacturing and service industries. Business service industries are subject to more specialization than non-business service industries while the manufacturing sector contains the lowest representation of industries with specialized places. Finally, we compare results for specialization with localization and show that both measures contribute to our understanding of industry and place specific agglomerative forces.

JEL Classification: R12

Keywords: industrial specialization; concentration metrics; agglomeration

Acknowledgements: We are grateful to Anthony Briant, Eric Brunner, Gilles Duranton, Christian Hilber and Stephen Ross, who provided invaluable feedback on earlier drafts of this paper. We appreciate the comments of two anonymous reviewers as well as the editor, Gilles Duranton. We thank seminar participants at the 2010 I Workshop on Urban Economics at the Institut d'Economia de Barcelona (IEB), the 2010 Skye Seminar on urban economics and real estate in Sleat, Scotland and the 2010 Boston Federal Reserve seminar on urban economics.

*University of North Carolina-Charlotte Email: stephen.billings@uncc.edu

†Quinnipiac University Email: Erik.Johnson2@quinnipiac.edu

1 Introduction

Industry agglomeration has been of interest to economists at least since [Marshall \(1920\)](#)'s discussion of the benefits to production for businesses that spatially concentrate. Modern examples include information technology firms in Silicon Valley, furniture producers in western North Carolina, and advertising firms in Manhattan. Surveys of the literature in [Rosenthal and Strange \(2003a\)](#) as well as throughout [Glaeser \(2010\)](#) highlight a sizeable number of agglomeration studies contributing to our understanding of both the scale, industrial scope and determinants of industrial concentration.

As interest in agglomeration has grown, empirically defining and measuring industry concentration has become more important. One can bisect measures of agglomeration into those that capture localization, defined by the overall concentration of specific industries across places, and specialization, defined as the concentration of an industry within a given place. Localization is the subject of recent research to develop new methodologies for measuring industrial concentration. [Ellison and Glaeser \(1997\)](#) derive a random utility model based index of localization that addresses the concern that perceived industry concentration may be a result of random clustering or the 'Dartboard Effect'. [Duranton and Overman \(2005\)](#) extend the literature by developing a nonparametric based measure of localization that overcomes the Modifiable Areal Unit Problem (MAUP) and provides a statistical test of significance.¹ Measures of specialization are less evolved methodologically and primarily limited to location quotients and Herfindahl based indices.² These ratio based metrics rank industrial concentration across places, but are subject to the same shortcomings addressed by [Ellison and Glaeser \(1997\)](#) and [Duranton and Overman \(2005\)](#) in the localization literature. Specifically, existing measures of specialization may be biased due to the MAUP and the 'Dartboard Effect' and are unable to provide a measure of statistical significance.

The current paper contributes to the agglomeration literature by introducing a new methodology to measure industrial specialization. This metric aims to improve our un-

¹MAUP relates to any form of statistical bias that occurs when data is aggregated to spatial units ([Openshaw \(1984\)](#)). In measuring industry concentration, results would be sensitive to both the size and shape of spatial units used for aggregation. Additionally, the boundaries of these spatial units may split up industry clusters thereby underestimating industry concentration.

²Empirical methods for measuring specialization are well summarized by [Holmes \(2004\)](#). Some highly cited empirical applications of these measures of specialization include [Krugman \(1991\)](#), who developed an index of regional specialization based on location quotients; [Glaeser \(1992\)](#) who uses location quotients to test the relationship between regional specialization and the growth of cities, and [Henderson et al. \(1995\)](#) who incorporates a Herfindahl based measure of diversity/specialization to test the role of Marshall-Arrow-Romer and Jacobian externalities in the concentration of manufacturing across U.S. cities.

derstanding of the composition and spatial pattern of industrial specialization within and across different resolutions of data and geographic areas. We formally introduce a statistical test for industrial specialization with three desirable properties. First, we address the MAUP by incorporating a bivariate kernel density estimator of industry concentration, which creates a continuous surface over the study area that is not defined by geographical units and boundaries. Second, we control for the ‘Dartboard Effect’ by first creating an empirical null distribution of establishment concentration based on a counterfactual of randomly located industries. We are then able to directly quantify industry concentration at a given place in the form of local p-values by comparing each place’s null distribution of establishment density to the density of a specific industry. Third, we provide a correction for multiple hypothesis testing in order to determine the statistical significance of specialization for a large number of places and industries.

We confirm that our methodology controls for the dartboard effect and is insensitive to the size of both industries and places through a Monte Carlo experiment. Subsequently, we apply our test for specialization to a single urban area, Denver-Boulder-Greeley CMSA, U.S. Results incorporate a different scale of analysis than has traditionally been incorporated into studies of specialization as well as tests the statistical power of our technique across rural, suburban and urban locations. Our application highlights some new stylized facts about the geographic pattern of industry clustering within a single urban area. The relationship between specialization and urbanization shows that places with greater underlying commercial density tend to specialize in more and different types of industries than their lower density counterparts. Across sectors, business services contain the largest portion of industries subject to specialization, while manufacturing contains the least. Empirical results highlight where specialization occurs for individual industries as well as the overall urbanization patterns for different sectors of the economy.

Additionally, we contrast industries subject to specialization with localized industries using the methodology of [Duranton and Overman \(2005\)](#). The relationship between specialization and localization can help disentangle the role of industry and place specific factors. To illustrate the difference between specialization and localization, consider the advertising industry, which is described as heavily concentrated in Manhattan by [Arzaghi and Henderson \(2008\)](#). This industry is localized since a large share of all U.S. advertising establishments are in Manhattan. At the same time, advertising concentration exceeds its share of general industry concentration in Manhattan indicating that Manhattan specializes in advertising. Therefore, the industry is both localized and subject to specialization. While specializa-

tion and localization often occur together, this relationship may not always be true. For example, our test for specialization finds three suburban clusters of establishments from the Offices of Physicians (NAICS 6211) industry, just north and southwest of downtown Denver. We identify these clusters as specialized places since they exceed the expected industrial concentration for these places, but this industry does not contain sufficient establishment concentration across all places in our study area for one to conclude localization. Our results support a large presence of specialization across industries with 62% of our 4 digit NAICS industries subject to specialization in at least one place while only 29% of all industries exhibit significant localization.

We continue with Section 2, where we describe our dataset and the range of industry categories we incorporate into our test for specialization. In Section 3, we detail our bivariate kernel density estimator of establishment concentration. In Section 4, we construct local p -values for each place and industry. We then adjust the critical values used to conclude specialization to correct for the fact that we have more than a single hypothesis test in Section 5. Section 6 identifies which industries are subject to specialization and Section 7 describes where specialization occurs. Section 8 discusses the relationship between industrial specialization and localization. Section 9 concludes.

2 Data

Implementing our test for specialization requires spatially disaggregated establishment level business data, for which we use the Quarterly Census of Employment and Wages (QCEW) Program (formerly known as ES-202) dataset. The QCEW is a cooperative program involving the Bureau of Labor Statistics (BLS) of the U.S. Department of Labor and State Employment Security Agencies. The QCEW program produces a comprehensive tabulation of employment and establishments for workers covered by state unemployment insurance laws.³ QCEW data provides the number of covered workers who worked during, or received pay for, the pay period including the 12th of the month.

We use establishment level data for QCEW covered firms for the 4th quarter of 2006 that are located in the Denver-Boulder-Greeley CMSA.⁴ This urban area contains 2.6 million people over 13,679 square kilometers. In Colorado, any business that paid wages of at least

³Excluded employees include members of the armed forces, the self-employed, proprietors, domestic workers, unpaid family workers, and railroad workers covered by the railroad unemployment insurance system.

⁴This CMSA includes eight counties: Adams, Arapahoe, Boulder, Broomfield, Denver, Douglas, Jefferson, and Weld.

\$1,500 in any quarter of the previous year, or employed at least one person for any part of a day for 20 weeks during the previous year must pay state unemployment insurance and thus is included in our dataset.

This data incorporates geographic information for the physical location of the establishments as well as mailing and corporate headquarters. Physical addresses are transformed into points with corresponding latitude and longitude coordinates by the QCEW Program.⁵ The dataset has a population of 79,038 establishments that represent industries across both the manufacturing and service sectors. In order to provide sufficient establishment representation in a given industry, we conduct our analysis at the 4 digit NAICS industry classification. We examine all industries with NAICS codes 3111 through 8142 (258 industries in our dataset).⁶ While the main focus of this paper is on identifying specialization at the 4-digit NAICS level, we group these industries into three main industry sectors for the purpose of exposition: NAICS 3111 through NAICS 3399 is classified as Manufacturing; NAICS 4231 through NAICS 4251 as well as NAICS 4811 through NAICS 6244 as Business Services; NAICS 4411 through NAICS 4543 as well as NAICS 7111 through NAICS 8142 as Non-Business Services. For these three industry sectors, Manufacturing contains 2,706 establishments, Business Services has 56,703 establishments and Non-Business Services includes 19,629 establishments.

3 Measuring Place Specific Establishment Concentration

The intuition behind our test for specialization may be illustrated through a comparison with the most commonly used measure of specialization, the location quotient (LQ). Typically, a location quotient is based on aggregate counts of establishments or employees at Census tract, county, or state geographies. For our discussion, we present the LQ as:

$$LQ_{i,j} = \frac{e_{i,j}/e_j}{e_i/e} \quad (1)$$

The numerator $e_{i,j}/e_j$ represents place i 's share of establishments (e) in industry j and

⁵Only 12.1% of establishments did not provide accurate enough geographic information to allow assignment of latitude and longitudes and are thus excluded from analysis. The excluded establishments are spread across industries and often excluded other data fields.

⁶These classifications include all manufacturing and service (wholesale trade, retail trade, transportation, and information through other services) industries and excludes agriculture, mining, utilities and construction.

the denominator e_i/e is place i 's share of total establishments across all industries. Location Quotients above one indicate above average specialization. Overall, the location quotient illustrates two properties that are necessary for correctly defining specialization. First, the measure must account for the likelihood that a randomly drawn establishment from a given industry would locate in a given place (the numerator). Second, the measure must control for the likelihood that a randomly drawn establishment from any industry would locate in this place (the denominator).

In measuring specialization across an urban area with a variety of commercial centers that vary in both size and density, any configuration of spatial units will likely violate both of [Arbia \(1989\)](#) and [Amrhein and Reynolds \(1997\)](#)'s data conditions⁷ necessary for no distortion due to the MAUP. Therefore, we begin our test for specialization by creating a measure of establishment concentration that controls for the MAUP. Specifically, we estimate a kernel density function based on the physical location of individual establishments.⁸ Kernel density functions generate a nonparametric and continuous measure of establishment location.

By adopting the kernel estimator, we generate a weighted average at a given location based on neighboring point intensity. Therefore, estimated establishment density is insensitive to the location of administrative or other geographic borders. This estimator may be interpreted as the probability that a randomly chosen establishment is found in a given location across the study area. If this probability is conditioned on industry, then the surface is comparable to the numerator of our LQ since both of these measures represent the establishment density of an industry at a given location.

To estimate a kernel density across our study area, we must choose both the kernel function and bandwidth. We base the kernel estimator on a bivariate Gaussian density function and use a smoothed cross validation (SCV) procedure to estimate the bandwidth.⁹ The kernel density estimator sums the values of the kernel functions generated at each establishment point and then divides by the total number of establishments in the sample.

⁷Condition 1 is the equivalence of spatial units in terms of size, shape and neighboring structure and condition 2 is the absence of spatial autocorrelation.

⁸The kernel smoothing of establishment location data has been justified in the agglomeration literature for several different reasons. Kernel smoothing may aid in overcoming data inaccuracies in establishment location due to measurement error ([McMillen and Klier \(2008\)](#) and [Duranton and Overman \(2005\)](#)) or the inexact nature of establishment location, where the actual location that an establishment selects may be proximate to its ideal location due to site availability. A kernel density estimator models point concentration as a decreasing function of distance to neighboring points, which is consistent with how the literature commonly models industry spillovers ([Duranton and Overman \(2005\)](#) and [Ellison et al. \(2010\)](#)).

⁹As discussed by [Waller and Gotway \(2004\)](#) and [Duong and Hazelton \(2005a\)](#), the choice of functional form for kernel estimation generates small differences in estimated densities, but the choice of bandwidth has significant consequences.

In our case, we define $\mathbf{x} = (x_1, x_2)$ as corresponding latitude and longitude coordinates for the N establishments in a given sample. Correspondingly, the set of points incorporated into the kernel estimator at each \mathbf{x} are given by $\mathbf{X}_\ell = (X_{\ell,1}, X_{\ell,2})$ for $\ell = 1, 2, \dots, N$. Together, the bivariate Gaussian function $K(\mathbf{x})$ and the 2×2 bandwidth matrix \mathbf{H} determine the shape of the kernel density estimator $\hat{f}(\mathbf{x}; \mathbf{H})$. The bivariate kernel density estimator is simply a two dimensional expansion of the commonly used univariate kernel density estimator, which we model through vector and matrix notation in Equation 2.

$$\hat{f}(\mathbf{x}; \mathbf{H}) = N^{-1} \sum_{\ell=1}^N |\mathbf{H}|^{-1/2} K(\mathbf{H}^{-1/2}(\mathbf{x} - \mathbf{X}_\ell)) \quad (2)$$

$$\mathbf{H} = \begin{pmatrix} h_1^2 & h_{1,2} \\ h_{2,1} & h_2^2 \end{pmatrix} \quad (3)$$

The choice of \mathbf{H} is debated in the literature and may have a significant impact on estimates of $\hat{f}(\mathbf{x}; \mathbf{H})$.¹⁰ In economic applications of kernel density estimation, scholars (e.g. [Duranton and Overman \(2005\)](#), [McMillen and Klier \(2008\)](#) and [Ellison et al. \(2010\)](#)) incorporate the *rule of thumb* bandwidth selection based on Section 3.4.2 of [Silverman \(1986\)](#). However, this type of bandwidth selection procedure may not be the best choice in our case because it assumes zero covariance and as discussed by [Wand and Jones \(1995\)](#) may oversmooth the data, thus masking the presence of multi-peaked surfaces. These two issues are relevant to our data and study area. First, off-diagonal elements of the bandwidth matrix should be non zero when establishments are aligned with physical features such as roads or rivers, which can be oriented in directions other than north-south or east-west. In our study area, a number of establishments are located on the northwest-southeast oriented US-36 corridor between Denver and Boulder. Second, since we incorporate downtown Denver as well as the surrounding areas such as Boulder and Greeley, and the Denver tech center, the study area is not single peaked.¹¹

Therefore, we estimate \mathbf{H} using the smoothed cross validation (SCV) technique introduced by [Hall et al. \(1992\)](#), which has been shown by [Duong and Hazelton \(2003\)](#) and [Duong and Hazelton \(2005b\)](#) to have a low Mean Integrated Square Error (MISE) for a range of target density shapes, an excellent convergence rate for small sample sizes, and an

¹⁰See [Wand and Jones \(1995\)](#), [Scott \(1992\)](#) and more recently [Duong and Hazelton \(2005a\)](#) and [Hall and Kang \(2005\)](#) for discussions on the different methodologies for selecting \mathbf{H} .

¹¹[Redfearn \(2007\)](#) and [McMillen \(2001\)](#) find the presence of multiple commercial centers of varying sizes within most large U.S. urban areas.

ability to accurately estimate the off-diagonal elements of the bandwidth matrix. The SCV bandwidth selection procedure is more formally discussed in the Appendix.

Once we determine the appropriate bandwidth for a given sample, $\widehat{\mathbf{H}}_j$, is substituted into Equation 2 to produce a kernel density estimate across our study area for industry j . Since the kernel smooths point data and is unbounded, the surface will necessarily cross the boundaries of the study area. We therefore impose a simple 2-dimensional modification of Silverman (1986)'s reflection method. This technique reflects smoothed data for values of (x_1, x_2) outside the study area back into the study area and then assigns zero values to (x_1, x_2) outside the study area.¹² Karunamuni and Alberts (2005) discuss the potential pitfalls of various reflection algorithms, but in practice any bias imposed by reflection on our bivariate kernel density estimation is minimal because of the large amount of undeveloped and sparsely developed land on the fringes of our study area. Reflection concerns are further mitigated by the fact that any statistical test using our kernel estimate involves a counterfactual generated from a kernel estimated with the same reflection method.

A graphical example of a bivariate kernel density estimate for the full population of all 79,038 establishments in the population is shown in Figure 1. The main area of establishment concentration is centered on downtown Denver and extends to secondary commercial centers in the south, west and northwest. The figure also highlights the discretization that we employed by creating a grid that encompassed our study area comprised of cells of approximately 5 square kms. The resulting grid of 51 by 51 cells is then overlaid onto our bivariate kernel density with the kernel density values assigned to the centroid of each grid cell and represent the density estimate for a given place. In order to verify that our bivariate kernel density estimator controls for MAUP, we formally applied our test for specialization to the case when we discretize the study area into larger 15 square km grid cells or incorporate Silverman (1986)'s *rule of thumb* bandwidth estimator and results find the same number and composition of four digit industries that are subject to specialization by at least one place.¹³

In the first stage of our test, we apply the kernel density and discretization algorithm to the population of establishments in each 4 digit industry. As one would expect, there is a large amount of variation in establishment density across industries. Figure 2 provides

¹²Specifically, for $\hat{f}(x_1 < \min(x_1), x_2)$, where x_1 is outside the study area, we assign its density to $(x_1 = (\min(x_1) + (\min(x_1) - x_1)), x_2)$ and replace \hat{f} with zero. Correspondingly, we assign density values to zero for locations where $(x_1, x_2 < \min(x_2))$ and the density is given to $(x_1, x_2 = (\min(x_2) + (\min(x_2) - x_2)))$. This process is replicated for all densities where $(x_1 > \max(x_1), x_2)$ and $(x_1, x_2 > \max(x_2))$ and densities are assigned to $(x_1 = (\max(x_1) - (x_1 - \max(x_1))), x_2)$ and $(x_1, x_2 = (\max(x_2) - (x_2 - \max(x_2))))$ correspondingly.

¹³The number of industries subject to specialization are unchanged using approximately 0.5 sq square km and 1 square km grid cells as well as shifts in the grid to the east and north.

examples for two specific industries, NAICS 5411 - Legal Services, and NAICS 5417 - Scientific Research & Development Services. It is clear from the industry population kernels that Legal Services contains multiple dense centers in downtown areas, while Scientific Research & Development Services contains lower density centered on Boulder and Denver.

4 Local P Values

The denominator of the LQ illustrates the need to scale our measure of industry specific concentration by a place's commercial density. Therefore, the next step in the construction of our estimator is to compare the kernel density estimate of a individual industry at a given place to what would potentially be observed from random draws of the population. We begin by identifying a population counterfactual based on randomly located industries and then estimating the full distribution of potential establishment concentrations across places for an industry. Finally, we compare the results to the actual industry concentrations to generate a base measure of statistical significance in the form of local p-values.

Similar to [Duranton and Overman \(2005\)](#), our sampling procedure to determine the counterfactual of randomly located industries has two specific criteria: 1) the sample should be drawn from the set of locations where a establishment could potentially locate, and 2) the sample size used in constructing the counterfactual must be equal to the number of establishments in the industry. Since our data contains two distinct types of establishments, manufacturing and service, we split our counterfactual accordingly.¹⁴ We assume that an establishment in a given service industry (NAICS 4000 to NAICS 8142) such as a grocery store or a dental office can reasonably locate in any service site. The same holds for specific manufacturing industries (NAICS 3000 to NAICS 3999) and all manufacturing sites. This strategy helps control for potential zoning regulations as well as other unobservable constraints on industrial location for manufacturing and service industries.¹⁵

For each industry, j , we construct a counterfactual based not only on potential establishment locations but also on an industry's establishment count, N_j .¹⁶ We then randomly select N_j locations from the set of all service or manufacturing establishment sites without

¹⁴Using all establishments as a counterfactual, we would expect more specialization because of the use of infeasible sites in determining the benchmark of random location.

¹⁵We are less concerned with establishment size constraints on our counterfactual given that [Duranton and Overman \(2008\)](#) find results unchanged when restricting large establishments to only large establishment sites.

¹⁶Restricting our counterfactual to the same number of establishments as our industry of interest accounts for any variation in the estimated density due to the sample size of the point process.

replacement. The resulting point data is then smoothed with the kernel function and densities are assigned to the corresponding grid cells. We apply the \mathbf{H}_j derived for each industry to that industry’s corresponding counterfactual in order to provide a consistent bandwidth when comparing industry densities to our counterfactual.¹⁷ We repeat this process of random point selection and kernel density estimation 50,000 times to create the empirical null distribution.

For each place $i = (1, \dots, I)$ in our study area, we compare the industry establishment density to the relevant empirical null distribution to create local p values (p_i^{local}). These values of p_i^{local} indicate the portion of our 50,000 kernel densities that are greater than the industry kernel density at place i .¹⁸

Our local p values represent the probability of falsely rejecting the null hypothesis of no specialization for each place and thus represent a single test for specialization at place i . For each industry j , each grid cell contains a unique $p_{i,j}^{local}$ based on the corresponding null distribution. This $p_{i,j}^{local}$ is a pivotal statistic, which allows one to compare results within a given industry across all places irrespective of the underlying heterogeneity in null distributions by location.¹⁹

Figure 3 displays local p-values for our example industries of Legal Services and Scientific Research & Development Services. We scale these p-values as $(1 - p_i^{local})$ to ease visual comparison to earlier population kernels. Therefore, areas of greater specialization relative to the empirical null distribution correspond to higher values on the z-axis. Corresponding contours for these p-values are provided in Figure 4. Darker areas in this figure indicate places of higher establishment concentration for an industry relative to randomly located industries. Legal Services are characterized by a multitude of higher and lower p-values across the study area, indicating that Legal Services tend to locate in multiple spatially concentrated clusters. These clusters follow population centers for this study area. For Scientific Research & Development Services, the bottom of Figure 4 shows corresponding p-value contours with relatively more concentration (darker areas) in/around Boulder and relatively less in Denver compared to the counterfactual. Referring to the population density

¹⁷We considered estimating \mathbf{H} uniquely for each estimated kernel (industry or counterfactual), but wanted to demonstrate that differences in kernel bandwidths between our industry and counterfactual kernels were not influencing our results.

¹⁸For example, if place i had a \hat{f}_i that exceeded or equaled 40,000 of our 50,000 kernel density estimates based on random point selection, then $p_i^{local} = 0.2$

¹⁹The main difference between Duranton and Overman (2005) local confidence bands and our local p-values is that we are scaling our local confidence bands to make comparisons easier across places. In essence, a p-value in one place represents the same probability of containing establishments for a randomly located industry as a p-value in any other place.

shown in Figure 2, we can see that the population kernel is captured in two peaks, one over Boulder and the other over Denver. The place specific local p-values highlight that kernel densities in Boulder and Denver for Scientific Research and Development Services are more likely to represent non-random clustering in Boulder than in Denver.

5 Global P-Values

Lower local p-values provide greater evidence that a place specializes in a given industry. However, if we are interested in evidence that an industry is specialized in *any* place, then inference based upon local p-values will overstate the amount of specialization. For example, assume that we define a standard critical local p-value of 0.05, and then perform hypothesis tests for an industry across all 2,601 places in our study area. Even if establishments were just randomly distributed across the study area, we would still expect to find 130 places where we reject the null hypothesis of no specialization for a given industry. This result would lead us to naively conclude that all industries are subject to specialization in multiple places. This issue has been termed the multiple hypothesis testing problem. Though well established in statistics and biostatistics, economists have only recently begun to recognize and properly correct for the flawed inference due to Type I errors under multiple hypothesis testing in empirical research. In recent work, [Romano and Wolf \(2005\)](#) stressed the need to minimize empirical data snooping for ‘false positives’ by controlling for familywise error rates (p^{FWE}). This process entails adjusting the critical values for each of the individual hypothesis tests to ensure that the probability of rejecting the null for *any one* of the multiple hypothesis tests is approximately equal to the p^{FWE} .²⁰ Therefore, we define a familywise error rate (here we choose 5%) and adjust the threshold (p^{adj}) for concluding statistical significance so that a false positive test for specialization only occurs in a prespecified percent of randomly located industries.

The Bonferroni correction is a classic and simple method for deriving the threshold value p^{adj} from a predetermined familywise error rate (p^{FWE}). This procedure divides the desired p^{FWE} by the number of hypothesis tests to find p^{adj} .²¹ For our study area, with 2,601 individual hypothesis tests, the Bonferroni correction defines $p^{adj} = 0.2 \times 10^{-4}$. This adjusted

²⁰Some examples of recent economics papers that adopt FWE corrections include [Anderson \(2005\)](#), [Kling et al. \(2007\)](#), [Bifulco et al. \(2008\)](#) and [Ross et al. \(2008\)](#)

²¹The logic behind this test is that each of the I places has a probability p^{adj} of being less than a critical value. The probability of *all* places being greater than p^{adj} is $(1 - p^{adj})^I$. For small levels of p^{adj} , one can approximate $p^{adj} = p^{FWE}/I$.

p-value is simple to compute, but it overcorrects if the hypothesis tests are correlated. For example, consider a case where all of the hypothesis tests were perfectly correlated, then the correct choice for p^{adj} would be exactly equal to p^{FWE} . In our case, where p-values are derived from smoothed and generally spatially correlated data, the Bonferroni method is too conservative and will underestimate specialization.

The failure of the Bonferroni method to account for correlation across hypothesis tests has resulted in a myriad of alternative strategies, ranging from parametric tests that explicitly define the nature of the correlation, random field methods which are based on the topological characteristics of Gaussian random variables, to nonparametric bootstrap and permutation techniques based upon the re-sampling procedures in [Westfall and Young \(1993\)](#).²² For the purpose of testing for specialization, the permutation based methods are best suited to control for spatial correlation without needing to make parametric assumptions on the shape of the empirical null distributions across places.

Our goal is to determine a critical value p^{adj} , which will result in a positive test for *any* place across the study area due to randomness only 5% of the time (the familywise error rate). An outline to create the correct p^{adj} is defined as follows and is based on a step down adjusted p-value method as discussed in [Westfall and Young \(1993\)](#) and [Farcomeni \(2008\)](#). First, we randomly sample $N_{\tilde{j}}$ establishments without replacement in order to generate a randomly located industry, which we term pseudo industry \tilde{j} . Next, we apply our kernel density estimator \hat{f} to our $N_{\tilde{j}}$ establishments and assign the estimated density to each place $i = 1, \dots, I$. We then construct our empirical null using 50,000 replications of the Pseudo-industries from the relevant counterfactual of $N_{\tilde{j}}$ random establishment sites.²³ The empirical null determines local $p_{i,\tilde{j}}^{local}$ for pseudo industry \tilde{j} . We then select $\bar{p}_{\tilde{j}} = \min_{1 \leq i \leq I} p_{i,\tilde{j}}^{local}$. This routine represents one pass of our algorithm.

We repeat this algorithm to generate 50,000 values of $\bar{p}_{\tilde{j}}$. Sorting these values generates a distribution of global p-values. In order to determine the p-value that satisfies the familywise error rate, we select the $\bar{p}_{\tilde{j}}$ where only 5% of the ranked $\bar{p}_{\tilde{j}}$ are smaller. The resulting global critical p-value is given by $p_{N_{\tilde{j}}}^{adj}$ and is determined uniquely for every possible industry size in our dataset.²⁴ $p_{N_{\tilde{j}}}^{adj}$ is significantly smaller than the 0.05 naively determined critical value that

²²Of these three strategies, [Nichols and Hayasaka \(2003\)](#) finds that the nonparametric re-sampling procedures outperformed a series of parametrically defined Bonferroni and random field based critical values using simulated data.

²³As the number of places increases, more replications are required to ensure that a sufficient number of decimal places can be captured for resulting p-values.

²⁴By construction, this algorithm verifies that $p_{N_{\tilde{j}}}^{adj}$ generates a Type I error rate equal to the familywise error rate (5%) in our dataset.

ignores the problem of multiple hypothesis testing and greater than the 0.2×10^{-4} defined by the Bonferroni correction. Figure 6 displays a scatterplot of $p_{N_j}^{adj}$ by the number of establishments in our industries.²⁵ As shown, industries with few establishments generate larger global p-values and as the number of establishments increase, the global p-values appear to converge towards 0.15×10^{-3} .²⁶ An inverse relationship between $p_{N_j}^{adj}$ and smaller industry sizes occurs because of larger estimated bandwidths for industries with few establishment as well as greater information in the tails of Pseudo-industry based null distributions for industries with a greater number of establishments.

To verify that our test adequately controls for industry size and is equally able to detect specialization in urban and rural locations, we conduct a Monte Carlo experiment upon randomly located (Pseudo) industries. Formally, we test if our globally adjusted p-value generates false positives that are correlated with industry size or commercial density. We use the fact that our $p_{N_j}^{adj}$ are set such that 5% of Pseudo-industries exhibit at least one specialized place in order to examine the distribution of specialized places for randomly located industries. We apply $p_{N_j}^{adj}$ to 5,000 Pseudo-industries for industry sizes of 5, 10, 100 and 500 establishments and record which places never experienced specialization.²⁷ In order to later test the statistical significance of Monte Carlo results, we repeat this experiment twenty times. We also categorized each place into quartiles based on the full population kernel density with Quartile 1 being the least dense (rural areas) and Quartile 4 being the most dense (urban areas).²⁸

Table 1 provides our results and reported values are based on the median count of our twenty Monte Carlo experiments. The first column indicates the number of non-specialized places in total for each industry size. Results show an even distribution of non-specialized places (with median values between 1,304 and 1,380) across industry sizes with no relationship between industry size and the number of non-specialized places.²⁹ The even distribution

²⁵Given the computational burden of calculating $p_{N_j}^{adj}$ and the fact that $p_{N_j}^{adj}$ converges for larger samples of establishments, we use random samples of 250 establishments for industries with more than 250 establishments. Figure 6 displays $p_{N_j}^{adj}$ by the full number of establishments in an industry.

²⁶The global p-values have relatively high variances across similar sized industries due to the influence of a given industry's spatial distribution on estimated bandwidths.

²⁷Due to the large computational burden in implementing this Monte Carlo experiment, we made some simplifying assumptions on the kernel bandwidth by adopting a product kernel based on the 'rule of thumb'.

²⁸These quartiles of density capture heterogeneity in empirical null distributions across places.

²⁹Given that $p_{N_j}^{adj}$ allows 250 Pseudo-industries to be subject to specialization in at least one place, one expects to find at least a few hundred specialized places. We find considerably more specialized places and thus fewer non-specialized places because of the spatial dependence in our measure of industry concentration. This result tells us that our 250 Pseudo-industries that are subject to specialization generate 1,200 to 1,300 specialized places or approximately 4.6 specialized places per industry. This occurs because we are merely

of results holds across quartiles with a range of between 319 and 359 non-specialized places by quartiles and sample sizes. We formally test these results using a [Kruskal and Wallis \(1952\)](#) test. This nonparametric test uses the ranking of the number of non-specialized places for each pseudo industry across industry sizes or quartiles to test the equality of population medians among the groups. The null hypothesis is that all groups are drawn from identical distributions. We implement this test separately across the four industry sizes and four quartiles within each industry size and report the results in [Table 1](#). The fact that none of our Kruskal-Wallis tests can reject the null hypothesis indicates that our test for specialization is insensitive to industry size as well as the population density of places.³⁰ These properties are important in order for us to compare our results for specialization across industries as well as between downtown, suburban and rural places.

[Figure 5](#) provides results of the global test for specialization for our two example industries. Areas of black indicate areas where $p_{i,j}^{local} < p_{N_j}^{adj}$ for place i in industry j . Areas of white indicate where we accept the null hypothesis of no specialization. Legal Services experiences three distinct clusters spread across the study area. These correspond closely with the cities of Denver, Boulder and Greeley. Scientific Research & Development Services exhibits significant specialization in multiple neighboring places northwest of Denver (along US-36 and in Boulder).

6 Industrial Composition of Specialization

[Table 2](#) presents the results of our test for specialization summarized across all four digit NAICS industries as well as just the manufacturing, business services, and non-business services sectors. We find that 62.0% of all industries contain *at least* one place with significant specialization, with business services containing the highest portion of industries subject to specialization at 71.5%. The manufacturing sector contained the fewest portion of industries with 41.0%, while 70.2% of non-business service industries were found to be specialized in at least one place. The large representation of business service industries is consistent with perceptions of technology and professional clusters like Route 128 in Boston and Research Triangle Park in North Carolina. The smaller representation for manufacturing industries is

controlling for the number of Pseudo-industries that are subject to specialization and not the number of specialized places with Pseudo-industries.

³⁰Again, the notion of multiple testing comes into play here, since we are performing 8 different hypothesis tests. In fact, using a basic Bonferroni correction we would expect to find a p value < 0.1 from the 8 tests 57% of the time.

consistent with our use of a single urban area, which limits such agglomeration benefits as labor matching/pooling or access to specialized inputs.

Table 2 also describes the number of distinct sets (groups) of specialized places. We find that 24.0% of all four digit industries are subject to specialization in more than one set of places and 9.7% of industries are in three or more sets of specialized places. Manufacturing is the least likely of the broad industry classifications to be subject to specialization, and when specialization occurs, it is concentrated in one portion of the urban area, with only 10.3% of all industries being in more than one set of specialized places. In contrast to Manufacturing, 24.6% of Non-Business Services and 32.5% of Business Services locate in more than one set of specialized places. The presence of approximately a quarter of industries containing multiple sets of specialized places in these two classifications suggests that agglomerative forces are present at varying scales across industries.

To further explore the industrial composition of specialization, Table 3 provides the percent of four digit industries with any specialized places for more detailed industry sectors. Focusing on column 1 highlights some variation in trends within the business and non-business service sectors. Eight of the nine four digit NAICS technology and professional industry classifications (NAICS 54) are subject to specialization. Other highly specialized sectors are Wholesale Trade, Real Estate & Rental & Leasing and Accommodation & Food Services. These industries are not commonly discussed in the agglomeration literature. Their specialization may be due to industry specific spillovers, but place specific amenities such as access to highways or consumer markets likely matter. Service industries often not subject to specialization include Educational Services and Arts, Entertainment & Recreation. These industries are highly consumer dependent and competition between establishments providing similar products and services likely weaken agglomerative forces.

The second column of Table 3 provides results incorporating employment weights. We implement the weighting procedure of [Duranton and Overman \(2005\)](#) which multiplies each establishment by its number of employees and then generates kernel density estimates. This procedure is applied to both the industry density as well as randomly selected Pseudo-industry establishments. For example, employment weighting for an establishment with 10 employees would implement our earlier methodology on ten observations of that establishment. Table 3 highlights a substantial decrease in the portion of all four digit industries subject to specialization and this holds for a number of two digit industry sectors. In fact, we find only half the total number of industries subject to specialization when we incorporate employee weights. This decrease in specialization with employment weighting is consistent

across most industries. The main exceptions are Manufacturing and Transportation & Warehousing, which find almost the same number of industries irrespective of weighting.

The large change in results for specialization using employment weights may be attributed to a number of factors. First, the size of establishments (number of employees) may correlate with specialized places. [Holmes and Stevens \(2002\)](#) finds a positive relationship between establishment size and industry concentration while [Guimaraes et al. \(2009\)](#) find this relationship to be negative. For our analysis, we find that unweighted results identify industries that average 23.1 employees per establishment while weighted results capture industries that average 39.0 employees per establishment. The difference in average employment between these two groups is significant at the 10% level.

Second, the variation in employees by establishment within and across industries impacts industry concentration as well as our null distribution counterfactual. Heterogeneity in employment across null distribution establishments generates larger confidence bands upon which to conclude statistically significant specialization. In essence, establishments with a large number of employees may be necessary for an industry to be subject to specialization under employee weighting. Additionally, we find that the standard deviation in employees per establishment within the same industry is 57.7 for industries subject to specialization in the unweighted results and 87.3 for the weighted results. Higher standard deviation and greater employees per establishment may be required to generate sufficient density in some places to conclude specialization under employee weighting. Given that our unweighted methodology has greater power to detect specialization by smaller establishments and across industry sectors, we focus on unweighted results in subsequent analysis.

7 Spatial Composition of Specialization

The spatial composition of specialized places for all industries is given in [Figure 7](#). Downtown Denver is in the center and contains the places with the greatest number of industries subject to specialization. Secondary commercial centers such as Boulder in the Northwest, the Denver Technology Center to the South and Greeley to the North also contain a number of industries subject to specialization. However, specialization is not confined to only the most dense urban areas. A number of specialized places extend radially along transportation corridors and a majority of places in our study area specialize in at least one industry. Overall, the presence and scope of specialization has a positive relationship with the location of commercial centers in this urban area. This pattern is statistically and economically

significant, since our Monte Carlo results predict that if establishments are randomly drawn from the population, the presence or absence of specialization should not correlate with the commercial density of a place.

The relationship between specialization and urbanization is described in Table 4, which provides the frequency of industries subject to specialization by quartiles of population density. We describe urbanization based on quartiles of kernel density estimates from our full population. The first two rows compare the distribution of places across quartiles for non-specialized and specialized places. Comparing the distributions of non-specialized places versus specialized places highlights a positive relationship between urbanization and specialization. Specialized places have a greater representation in the densest quartile while non-specialized places have greater representation in the least dense quartile. As shown by the Monte Carlo experiment for Pseudo-industries, the ratio of non-specialized places found in the most dense to the least dense locations is between .92 and .99, while with our actual industry data, this ratio is 0.3. This pattern also holds true for the actual counts of specialized places. Overall, 26% of all places in the study area do not specialize in any industry and 43% of places specialize in only one or two industries.

The final column of Table 4 indicates that the bottom density quartile is more likely to not specialize or to specialize in only one industry and places that specialize in two or more industries have a greater representation in the top density quartile. From the most dense quartile, one sees that places that specialize in a large number of industries are almost exclusive to more urbanized places. Supporting a larger number of industries likely requires a sufficient concentration of commercial activity.³¹ For places in the bottom density quartile, 39% of places do not specialize in any industry and 47% of places specialize in only one industry. Correspondingly, only 14% of places in the top density quartile are not specialized in any industry. The second quartile finds 27% and the third quartile finds 24% of their places to not specialize in any industry. The second and third quartiles represents a number of suburban places and contain the greatest concentration of places that specialize in between one and four industries. Suburban places appear well suited to specialize in a few industries.

The relationship between specialization and urbanization may also vary by industry sector. We provide the number of specialized places for each industry aggregated to 2 digit industry sectors and the distribution of these places across quartiles of population density in

³¹One may be concerned that a high commercial density is required for us to conclude that a place has a large number of specialized industries. This concern is not due to the nature of our test for specialization, where concluding a place to specialize in one industry does not influence the test for specialization in another industry.

Table 5.³² Across industry sectors, places specializing in industries in the Finance & Insurance and Professional & Scientific & Technical Services are more often located in the densest places, while Accommodation & Food Services and Arts, Entertainment & Recreation have the smallest presence in the most dense places. Places in denser commercial areas almost never specialize in Arts, Entertainment & Recreation industries and this industry sector is heavily specialized in suburban or secondary commercial centers (61% of the specialization in this sector occurs in the third quartile). Industries in Manufacturing, Transportation & Warehousing, Information, Administrative & Waste Services and Accommodation & Food Services have a number of specialized places across quartiles and highlight a strong presence in both downtown as well as suburban places. Information and Manufacturing contain some higher technology industries and tend to locate along suburban transportation corridors and industrial parks between Boulder and Denver as well as south of Denver in the Denver Technology Center.

We next examine the relationship between specialization and urbanization for specific four digit industries. For ease of exposition and comparisons across four digit industries, we assign each specialized place a ranking based on its population density. We rank places from least to most dense and compute summary statistics based on these rankings across all specialized places within a given four digit industry.³³ We provide results for the industries located in the least dense, most dense and highest variance in density specialized places. We also identify the number of distinct sets of specialized places for each industry, where we define a set of specialized places as a single grouping of contiguous places.³⁴

In the top panel of Table 6, we identify the top ten industries based on specialized places in the least dense locations. One of these industries is Lawn & Garden Equipment & Supplies Stores, which is shown in the top panel of Figure 8. The bottom panel of this figure plots the rank of specialized places on the x-axis and their corresponding population density on the y-axis. This figure displays a dot for each specialized place in this four digit industry and highlights the presence of a number of specialized places in the low density areas. Low density specialization also occurs for Recreational Vehicle Parks & Recreational Camps, which locate in/around national wilderness areas in the foothills of the Rocky Mountains on the western edge of our study area. Some residential based industries on this list are Elementary & Secondary Schools and Agents & Managers for Public Figures. The former includes private

³²Percentages in columns are based on the number of specialized places for each 2 digit industry sector.

³³The least dense place is given a rank of 1 and the most dense place a rank of 2,601.

³⁴For example, Figure 5 shows that NAICS 5411 contains 3 distinct sets varying in size from a single place to a set of 15 contiguous places.

education institutions and the latter incorporates a number of home businesses. Other Support Services includes a number of industrial based business support services that require large tracts of land and likely serve businesses across the urban area. A number of four digit manufacturing industries such as Waste Collection and Agricultural Chemical Manufacturing often locate away from denser commercial areas due to negative production externalities.

The second panel of Table 6 shows the top 10 industries based on specialized places in the most dense locations. All ten industries are specialized in a single part of the urban area as given by the presence of only one set of specialized places. Figure 9 highlights the Advertising & Related Services, which contains a set of specialized places in downtown Denver.³⁵ Two main types of industries populate this list. First, Other Investment Pools & Funds and Professional, Scientific & Technical Services represent the large presence of financial and professional services in downtown Denver, which serve a number of central city businesses. Second, Social Advocacy; Business, Professional & Labor Organization and Grant making & Giving Services locate in order to access officials in the state capital, which is located in downtown Denver. The first group of industries likely benefit from both own industry as well as other industry concentration while the second group concentrates to access the state capital.

As shown in the bottom panel of Table 6, variance in the population density of specialized places occurs for a range of industries and is exclusive to industries with multiple sets of specialized places. The mix of denser and sparser specialized places highlight that specialization is not exclusive to certain commercial densities for these industries. Figure 10 displays Management of Companies and Enterprises, which captures establishments that contain equity interest in companies and may serve as administrative or corporate offices. This figure displays four distinct sets of specialization with two sets in higher density places and two sets in lower density places. These establishments tend to locate in both downtown and suburban areas. Waste Collection serves both residential and business locations. Other Personal Services, which include such diverse activities as bail bonding, parking lot/garages as well as dating services, generates a range of specialization and commercial densities.³⁶ The two air transportation industries are located proximate to airports and also along interstates and thus contain a mix of suburban locations.

³⁵The results are consistent with [Arzaghi and Henderson \(2008\)](#)'s discussion of advertising agency concentration in Manhattan.

³⁶In essence, some of the multiple sets of specialized places may just be a result of industry classification. This highlights a problem that is endemic to this literature, the mismatch between industry classifications (e.g. NAICS, SIC) and industry categorizations that best capture agglomerative forces.

8 Comparison of Agglomeration Measures

The relationship between specialized places for a specific industry and that industry’s overall degree of localization may highlight the role of industry specific as well as place specific agglomerative forces. Therefore, we implement a test for localization in order to highlight the relationship between specialization and localization across each industry. [Duranton and Overman \(2005\)](#) provide a well established test for localization that incorporates similar characteristics as our test for specialization. We begin by replicating the [Duranton and Overman \(2005\)](#) methodology using our Colorado dataset with our full set of industry classifications and note any modifications in our application.

8.1 Duranton & Overman Test for Localization

The first step in implementing the [Duranton and Overman \(2005\)](#) test for localization is to estimate a univariate kernel density based on $\frac{n*(n-1)}{2}$ unique pairwise Euclidean distances for all n establishments in a given industry.³⁷ This kernel may be defined for areas where the pairwise distance is less than zero, so data reflection is done following the [Silverman \(1986\)](#) technique. Kernel bandwidths are set along one dimension, the pairwise distance, using the [Silverman \(1986\)](#) ‘rule of thumb’ procedure. The counterfactual of randomly located industries is based on randomly sampling from all manufacturing or service establishments analogous to the methodology described in Section 4. We simulate a full empirical null distribution of kernel smoothed pairwise distances using 2,000 replications. Finally, local critical values are determined from the empirical null distribution for all possible pairwise distances.

The [Duranton and Overman \(2005\)](#) solution to the multiple testing problem is to create global confidence bands based on null distribution kernels. [Duranton and Overman \(2005\)](#) begin by sorting kernels at each pairwise distance such that 95% of the kernels lie entirely below the upper confidence band. The envelope of kernel density values that satisfy these criteria provide the global confidence band for each pairwise distance. These global confidence bands are conceptually similar to our globally adjusted p-value because they dictate that 95% of randomly located industries accept the null hypothesis of no localization at any pairwise distance.³⁸ We calculate the median pairwise distance in our dataset (25.6 km)

³⁷To avoid computational problems for large industries, we randomly draw a subset of establishments equal to 200 for any industry or counterfactual with more than 200 establishments.

³⁸The main applied difference is that we generate local p-values so that a single p-value provides a global critical value instead of the Duranton & Overman global confidence bands, which vary across pairwise

and conclude localization when an industry specific kernel exceeds the global upper confidence band for any distance less than or equal to 25.6 km.³⁹ Correspondingly, we conclude dispersion for distances greater than 25.6 km.⁴⁰

The graphical results of the [Duranton and Overman \(2005\)](#) test for localization in our two sample industries are shown in [Figure 11](#). For Legal Services, the test concludes localization because the industry kernel exceeds the global upper confidence band given by the dotted line for all distances less than approximately 10km. A second significant range of distances occurs from 35 to 40 km, and represents the distance between clusters of establishments. In other words, the first peak defines the intensity of existing clusters and the second peak represents the distance between the clusters. In the lower panel of [Figure 11](#), the industry based kernel for Scientific Research & Development Services exceeds the global confidence band at distances of 35 to 55 km, which is in a range greater than the median pairwise distance of 25.6km. Therefore, no localization is concluded for this industry.

We provide the results for four specific industries to highlight the range of possible outcomes for the 258 industries upon which we test for specialization and localization. Each set of figures provides an industry’s population kernel, globally significant specialized places, plots of specialization by commercial density and the Duranton & Overman estimate of localization for comparison. [Figure 12](#) provides results for NAICS - 3118 Bakeries & Tortilla Manufacturing. Comparing our results with Duranton & Overman show that this localized industry contains one specialized place. Panel (c) of [Figure 12](#) shows that this specialized place is located in a medium density commercial center just east of Downtown Denver. An example of an industry with multiple sets of specialized places is given by NAICS 5171 - Wired Telecommunications Carriers in [Figure 13](#). Results emphasize the presence of two sets of specialized places in denser portions of the study area. According to Duranton & Overman, this industry would be characterized as localized at two different scales, less than 4 km and also between 19 and 22 km. [Figure 14](#) provides results for NAICS 4841 - General Freight Trucking and indicates the presence of multiple sets of specialized places in sparser locations along Interstates 25, 70 and 76. Duranton & Overman’s test would not find this

distances.

³⁹The maximum pairwise distance in our dataset occurs around 120 km.

⁴⁰Our results for localization are based on a small modification of Duranton & Overman, who only look at pairwise distances less than or equal to the median pairwise distance and use an upper and lower confidence band. In Duranton & Overman, pairwise distances that exceed the upper global confidence are concluded as localized and distances that fall below the lower global confidence threshold are designated dispersion. Both methods should provide similar results since the kernel density integrates to one over the full range of pairwise distances. We make this modification because we later test the sensitivity of concluding localization for distances other than the median distance in our dataset.

industry to be localized. An industry that provides neither specialized places nor localization is NAICS 4422 - Home Furnishing, which is given in Figure 15.

8.2 Specialization and Localization

Table 7 directly compares our estimates of specialization with those of localization by two digit industry sector. For the set of all industries, the differences are substantial. We find that 62.0% of all industries are subject to specialization in at least one place while only 28.7% of industries are localized.⁴¹ This difference is maintained for most two digit industry sectors. In Duranton & Overman’s study of national level U.K. manufacturing, they found that 52% of industries are localized. Using Duranton & Overman’s measure we find only 15% of manufacturing industries are localized. The smaller estimate for manufacturing in our study area relative to the U.K. dataset is consistent with the scale of our dataset being limited to a single urban area. This scale of analysis limits such localization benefits as labor matching/pooling and access to specialized inputs.

Eighteen of the nineteen Wholesale Trade industries (94.7%) are subject to specialization, while only 52.6% of Wholesale Trade industries are localized. This difference may be attributed to specialized places with low establishment concentrations due to Wholesale Trade industries locating away from traditional downtown commercial centers given their large scale operations and lack of walk-in traffic. In Retail Trade, 74.1% of four digit industries are in at least one specialized place while only 25.9% are localized. Large retail and strip malls may generate specialization, but likely do not represent enough overall industry concentration to conclude localization. One of the most striking differences between industry findings for specialization and localization is in Professional, Scientific & Management Services where eight of nine industries are in specialized places, but only four are considered localized. These higher technology industries are located in downtown Denver, secondary commercial centers in Boulder, Golden and along US-36 connecting Denver to Boulder as well as in the Denver Technology Center in the southern portion of the study area. Specialization in multiple portions of the urban area may be too spread out to conclude localization. Accommodation & Food Services industries locate across the study area and contain the largest difference between industries subject to specialization and industries found to be localized.

Graphical and industry results highlight the presence of multiple distinct clusters of

⁴¹The substantially fewer localized industries than industries subject to specialization is minimally influenced by the pairwise distance used to classify localization and dispersion. For example, our finding that 28.7% of industries are localized increases to 33.3% and 40.3% if one adopts the 75th percentile of pairwise distances (40.5 km) and the 95th percentile of pairwise distances (69 km) for concluding localization.

establishments for a number of industries. Multiple clusters may not generate sufficient density of small pairwise distances to test positive for localization.⁴² Table 8 provides the number of industries by sets of specialized places for localized and non-localized industries. This table includes just the 159 industries subject to specialization in our dataset.⁴³ Our results show an even distribution of industries across the number of sets of specialized places for both localized and non-localized industries. Therefore, the difference between localization and specialization is not a function of multiple cluster industries.

The large number of non-localized industries subject to specialization is likely a function of the differing counterfactual used by these measures. We test the role of our specialization counterfactual, which is determined by a place's population density, by further comparing our 87 non-localized industries and 72 localized industries. Table 8 shows that specialized places contain a larger percentage of a localized industry's establishments than in the case of a non-localized industry. Specialized places contain only 16.4% of all establishments in non-localized industries, but 30.6% of all establishments in localized industries.

Additionally, we find that places specializing in localized industries generate an average density place rank of 2,420 or in the 7% most dense places for our study area. Of the 87 non-localized industries, the average density place rank is 2,133 or in the 18% most dense places. Implementing a bootstrapped t-test for differences in mean population kernel density between these two groups finds that they are statistically different from one another ($t = 5.74$).⁴⁴ These results show that our test for specialization detects clustering in a number of low density places that do not contribute enough to overall industry concentration to conclude localization for those industries. Furthermore, if one looks at the top ten industries by population density of specialized places in Table 6, nine of these industries are identified as localized. In contrast, only two industries are localized for the ten industries with the lowest density for specialized places. These results support the role of counterfactuals in differences between specialization and localization.

⁴²The primary concern is that the distance between clusters may generate a sufficient enough density in the Duranton & Overman measure of localization to conclude insignificant localization or even dispersion.

⁴³Only two industries tested positive for localization, but not for specialization.

⁴⁴The average population density for places specializing in localized industries is 0.00141 and for places specializing in non-localized industries is 0.00056.

9 Conclusion

In this study, we develop a new statistical test for specialization that is able to identify not only where specialization occurs, but also which industries are subject to specialization. We implement our test by constructing a bivariate kernel density estimator of establishment concentration within a given industry. We use establishment density estimates and a permutation based empirical null distribution of randomly located industries to assign probabilities of non-random clustering across places in our study area. Our technique derives a new global estimator for significant departures from randomness that accounts for spatial dependence across hypothesis tests and is unbiased for small samples. This methodology yields a measure of specialization that can be applied to econometric studies of agglomeration, yet still allow for statistical tests of the significance of specialization and controlling for the Modifiable Areal Unit Problem (MAUP).

Results indicate a positive relationship between urbanization and specialization with dense commercial places more likely to specialize and to do so in more industries. Industry results show that 62% of all industries contain specialized places and a quarter of these industries contain multiple sets of specialized places within a single urban area. When examining the relationship between specialization and localization, we show that a number of industries are specialized but not localized and these specialized places occur in sparser commercial areas. By industry sectors, results indicate the presence of suburban industry concentration in non-localized industries. This highlights the concern that simply identifying an industry as localized may miss meaningful agglomeration that aids in our understanding of specialized places like Silicon Valley.

Future research points toward econometric studies to isolate the benefits due to industry specific external economies from place specific amenities using this new test for specialization. Econometrically, studies have incorporated the localization measures of [Duranton and Overman \(2005\)](#) and [Ellison and Glaeser \(1997\)](#), but given the nature of localization, analysis is restricted to industry level observations (see [Ellison and Glaeser \(1999\)](#) and [Ellison et al. \(2010\)](#)). A number of papers on the determinants of agglomeration measure clustering based on counts of proximate establishments or employment.⁴⁵ Without a formal test to rule out random location, factors that influence overall industrial concentration may confound estimates. By detecting specialized places, subsequent research into the determinants of agglomeration can highlight the role of both industry and place specific factors.

⁴⁵For examples of this research, see [Arzaghi and Henderson \(2008\)](#), [Rosenthal and Strange \(2003b\)](#) and [Holmes \(1999\)](#).

References

- Amrhein, C. and Reynolds, H.: 1997, Using the getis statistic to explore aggregation effects in metropolitan toronto census data, *The Canadian Geographer* **41**, 137149.
- Anderson, M.: 2005, Multiple inference and gender differences in the effects of early intervention: A reevaluation of the abecedarian, perry preschool, and early training projects, *Journal of the American Statistical Association* **103**, 1481–1495.
- Arbia, G.: 1989, *Spatial Data Configuration in Statistical Analysis of Regional Economic and Related Problems*, Kluwer.
- Arzaghi, M. and Henderson, J. V.: 2008, Networking off Madison Avenue, *Review of Economic Studies* **75**(4), 1011–1038.
- Bifulco, R., Fletcher, J. M. and Ross, S. L.: 2008, The effect of classmate characteristics on individual outcomes: Evidence from the add health, *University of Connecticut Department of Economics Working Paper* .
- Duong, T. and Hazelton, M.: 2003, Plug-in Bandwidth Matrices for Bivariate Kernel Density Estimation, *Journal of Nonparametric Statistics* **15**, 17–30.
- Duong, T. and Hazelton, M.: 2005a, Convergence Rates for Unconstrained Bandwidth Matrix Selectors in Multivariate Kernel Density Estimation, *Journal of Multivariate Analysis* **93**, 417–433.
- Duong, T. and Hazelton, M.: 2005b, Cross-Validation Bandwidth Matrices for Multivariate Kernel Density Estimation, *Scandinavian Journal of Statistics* **32**(), 485–506.
- Duranton, G. and Overman, H. G.: 2005, Testing for localisation using micro-geographic data, *Review of Economic Studies* **72**(4), 1077–1106.
- Duranton, G. and Overman, H. G.: 2008, Exploring the detailed location patterns of UK manufacturing industries using microgeographic data, *Journal of Regional Science* **48**(1), 213–243.
- Ellison, G. D., Glaeser, E. L. and Kerr, W. R.: 2010, What Causes Industry Agglomeration? Evidence from Coagglomeration Patterns, *The American Economic Review* **100**(3), 1195–1213.
- Ellison, G. and Glaeser, E.: 1999, The determinants of geographic concentration, *American Economic Review Papers and Proceedings* **89**(2), 311–316.
- Ellison, G. and Glaeser, E. L.: 1997, Geographic concentration in u.s. manufacturing industries: A dartboard approach, *Journal of Political Economy* **105**(5), 889–927.

- Farcomeni, A.: 2008, A review of modern multiple hypothesis testing, with particular attention to the false discovery proportion, *Statistical Methods in Medical Research* **17**, 347–388.
- Glaeser, E. L.: 1992, Growth in cities, *Journal of Political Economy* **100**(6), 1126–1152.
- Glaeser, Edward (Ed.): 2010, *Agglomeration Economies*, University of Chicago Press, NBER.
- Guimaraes, P., Figueiredo, O. and Woodward, D.: 2009, Localization economies and establishment size: was Marshall right after all?, *Journal of Economic Geography* **9**(6), 853–868.
- Hall, P. and Kang, K.: 2005, Bandwidth Choice for Nonparametric Classification, *The Annals of Statistics* **33**(), 284–306.
- Hall, P., Marron, J. and Park, B.: 1992, Smoothed Cross-Validation, *Probability Theory and Related Fields* **92**, 1–20.
- Henderson, V., Kuncoro, A. and Turner, M.: 1995, Industrial development in cities, *Journal of Political Economy* **103**(5), 1067–1090.
- Holmes, T. J.: 1999, Localization of industry and vertical disintegration, *Review of Economics and Statistics* **81**(2), 314–325.
- Holmes, T. J.: 2004, Spatial distribution of economics activities in north america, *Handbook of Urban and Regional Economics Vol.4* .
- Holmes, T. J. and Stevens, J. J.: 2002, Geographic concentration and establishment scale, *Review of Economics and Statistics* **84**(4), 682–690.
- Karunamuni, R. and Alberts, T.: 2005, On boundary correction in kernel density estimation, *Statistical Methodology* **2**(3), 191–212.
- Kling, J. R., Liebman, J. B. and Katz, L. F.: 2007, Experimental analysis of neighborhood effects, *Econometrica* **75**(1), 83–119.
- Krugman, P.: 1991, *Geography and Trade*, MIT Press, Cambridge.
- Kruskal, W. H. and Wallis, W. A.: 1952, Use of ranks in one-criterion variance analysis, *Journal of the American Statistical Association* **47**, 583621.
- Marshall, A.: 1920, *Principles of Economics*, London: MacMillan.
- McMillen, D. and Klier, T.: 2008, Evolving Agglomeration in the U.S. Auto Supplier Industry, *Journal of Regional Science* **48**, 245–267.
- McMillen, D. P.: 2001, Nonparametric Employment Center Identification, *Journal of Urban Economics* **50**, 448–473.

- Nichols, T. and Hayasaka, S.: 2003, Controlling the Familywise Error Rate in Functional Neuroimaging: A Comparative Review, *Statistical Methods in Medical Research* **12**(), 419–446.
- Openshaw, S.: 1984, *'The modifiable areal unit problem'*, *Concepts and Techniques in Modern Geography*, Geo Books.
- Redfearn, C.: 2007, The topography of metropolitan employment: Identifying centers of employment in a polycentric urban area, *Journal of Urban Economics* **61**, 519–541.
- Romano, J. P. and Wolf, M.: 2005, Stepwise multiple testing as formalized data snooping, *Econometrica* **73**(4), 1237–1282.
- Rosenthal, S. S. and Strange, W. C.: 2003a, *Evidence on the Nature and Sources of Agglomeration Economies*, V. Henderson and J.F. Thisse (Ed.) Handbook of Regional and Urban Economics IV, Elsevier, pp. 2120–2149.
- Rosenthal, S. S. and Strange, W. C.: 2003b, Geography, industrial organization, and agglomeration, *Review of Economics and Statistics* **85**(2), 377–393.
- Ross, S. L., Turner, M. A., Godfrey, E. and Smith, R. R.: 2008, Mortgage lending in chicago and los angeles: A paired test study of the pre-application process, *Journal of Urban Economics* **63**, 902–919.
- Scott, D.: 1992, *Multivariate Density Estimation: Theory, Practice, and Visualization*, New York: John Wiley & Sons.
- Silverman, B.: 1986, *Density Estimation for Statistics and Data Analysis*, London: Chapman and Hall.
- Waller, L. A. and Gotway, C. A.: 2004, *Applied Spatial Statistics for Public Health*, Wiley Series in Probability and Mathematical Sciences.
- Wand, M. and Jones, M.: 1995, *Kernel Smoothing*, Boca Raton, FL: Chapman & Hall/CRC.
- Westfall, P. H. and Young, S. S.: 1993, *Resampling-Based Multiple Testing: Examples and Methods of p-Value Adjustment*, Wiley Series in Probability and Mathematical Sciences.

A Appendix: Smoothed Cross Validation Bandwidth Estimator

The smoothed cross validation bandwidth estimator represents a modification of a least squares cross validation (LSCV) technique. Therefore, we begin with describing the standard LSCV, which is given by,

$$LSCV(\mathbf{H}) = \int_{\mathbb{R}^2} \hat{f}(\mathbf{x})^2 dx - 2n^{-1} \sum_{\ell=1}^N \hat{f}_{-\ell}(\mathbf{x}) \quad (\text{A.1})$$

$$\hat{f}_{-\ell}(\mathbf{x}) = (N-1)^{-1} \sum_{\ell \neq k}^N |\mathbf{H}|^{-1/2} K(\mathbf{H}^{-1/2}(\mathbf{x} - \mathbf{X}_k)) \quad (\text{A.2})$$

This technique involves estimating \mathbf{H} based on minimizing $LSCV(\mathbf{H})$, which directly estimates Mean Integrated Squared Error (MISE), using a leave-one-out estimator ($\hat{f}_{-\ell}$). This technique is extended to smoothed cross validation (SCV) by pre-transforming the data in order to allow better estimation of \mathbf{H} under the large sampling fluctuations in estimates that often occur using standard cross validation techniques.⁴⁶ Specifically, we estimate an unconstrained version of SCV with a pre-sphering data transformation. These attributes are shown to improve kernel density estimation even with non-coordinate alignments of point patterns (?). The pre-sphering transforms the original data $\mathbf{X}_1, \mathbf{X}_2$ to $\mathbf{X}_1^*, \mathbf{X}_2^*$ by

$$\mathbf{X}^* = \mathbf{S}^{-1/2} \mathbf{X}$$

where \mathbf{S} indicates a full covariance matrix of the untransformed data. The optimal bandwidth \mathbf{H} is determined for each industry by minimizing the following expression using the transformed data \mathbf{X}^* .

$$\underset{\mathbf{H}}{\operatorname{argmin}} \quad SCV(\mathbf{H}) = \int_{\mathbb{R}^2} \hat{f}(\mathbf{x}^{*2}) dx - 2n^{-1} \sum_{\ell=1}^N \hat{f}_{-\ell}(\mathbf{x}^*) \quad (\text{A.3})$$

⁴⁶See [Hall et al. \(1992\)](#) for more details.

Table 1: Non-Specialized Places for Pseudo-industries

Industry Size (N)	All Places	Least Dense \rightarrow Most Dense				Kruskal-Wallis Test ($\chi^2_{(3)}$)
		1st	2nd	3rd	4th	
5	1,380	349	359	344	328	6.7 (0.08)
10	1,370	350	355	341	324	3.4 (0.34)
100	1,304	331	327	319	327	1.6 (0.66)
500	1,377	359	336	352	330	4.2 (0.24)
Kruskal-Wallis Test ($\chi^2_{(3)}$)	4.5 (0.22)	1.2 (0.75)	6.4 (0.10)	3.6 (0.31)	1.3 (0.74)	

Based on $p_{N_j}^{adj}$, 5% of Pseudo-industries contain at least one specialized place. p-values in parenthesis.

Each row provides the results of our test for specialization on 5,000 Pseudo-industries of size N_j and each cell indicates the median count of non-specialized places based on 20 repetitions of these 5,000 Pseudo-industries.

1st \rightarrow 4th indicate quartiles of the full population kernel density.

Table 2: Specialization for 4 Digit NAICS Industries

	All Industries	Manufacturing	Business Services	Non-Business Services
Percent of 4 digit NAICS Industries with any Specialized Places	62.0%	41.0%	71.5%	70.2%
Percent of 4 digit NAICS Industries with Multiple Sets of Specialized Places	24.0%	10.3%	32.5%	24.6%
Percent of Industries by Number of Distinct Sets of Specialized Places				
Zero	38.0%	59.0%	28.5%	29.8%
One	38.0%	30.8%	39.0%	45.6%
Two	14.3%	7.7%	21.1%	8.8%
Three	6.2%	2.6%	4.9%	14.0%
Four or More	3.5%	0.0%	6.5%	1.8%
Number of Industries	258	78	123	57

We define a distinct set of specialized places as a unique grouping of contiguous specialized places.

For example, Figure 5 shows that NAICS 5411 contains 3 distinct sets varying in size from a single place to a set of 15 contiguous places.

Table 3: Specialization

NAICS Code	Industry Name	No. 4 digit Industries	% of 4 digit Industries Subject to Specialization No Weights	% of 4 digit Industries Subject to Specialization with Employee Weights
31-33	Manufacturing	78	41.0%	39.0%
42	Wholesale Trade	19	94.7%	42.1%
44-45	Retail Trade	27	74.1%	3.7%
48-49	Transportation and Warehousing	23	69.6%	73.9%
51	Information	16	56.3%	37.5%
52	Finance & Insurance	11	72.7%	18.2%
53	Real Estate & Rental & Leasing	8	75.0%	12.5%
54	Professional, Scientific & Technical Services	9	88.9%	44.4%
55	Management of Companies & Enterprises	1	100.0%	0.0%
56	Administrative, Waste & Remediation Services	11	63.6%	36.4%
61	Educational Services	7	42.9%	0.0%
62	Health Care & Social Assistance	18	66.7%	22.2%
71	Arts, Entertainment & Recreation	9	55.6%	0.0%
72	Accommodation & Food Services	7	85.7%	28.6%
81	Other Services (except Public Administration)	14	64.3%	21.4%
	Total	258	62.0%	31.9%

Table 4: Specialization by Urbanization

# Industries Subject to Specialization	All Places	Least Dense → Most Dense				4th / 1st
		1st	2nd	3rd	4th	
0	678	256	178	156	88	0.3
> 0	1,923	394	472	494	562	1.4
0	678	256	178	156	88	0.3
1	760	306	217	136	101	0.3
2	371	55	89	113	114	2.1
3	202	18	31	75	78	4.3
4	149	10	56	47	36	3.6
5	85	2	24	32	27	13.5
6- 10	245	3	54	72	116	38.7
11- 15	48	0	1	11	36	-
16- 20	32	0	0	4	28	-
21- 25	14	0	0	2	12	-
26- 30	11	0	0	2	9	-
31- 49	6	0	0	0	6	-
50- 258	0	0	0	0	0	-
Total Places	2,601	650	650	650	651	

The final column provides the ratio of the number of places in the 4th quartile to the number of places in the 1st quartile.

Table 5: Specialization by Urbanization for Industry Sectors

NAICS Code	Industry Sector	# Specialized Places	Least Dense → Most Dense			
			1st	2nd	3rd	4th
31-33	Manufacturing	1,563	8%	22%	28%	41%
42	Wholesale Trade	980	2%	10%	23%	65%
44-45	Retail Trade	350	0%	4%	18%	78%
48-49	Transportation & Warehousing	1,547	6%	27%	33%	34%
51	Information	323	12%	20%	13%	55%
52	Finance & Insurance	188	1 %	1 %	7 %	92 %
53	Real Estate & Rental & Leasing	183	1%	13%	31%	55%
54	Professional, Scientific, & Technical Services	232	0%	1%	14%	85%
55	Management of Companies & Enterprises	20	15%	10%	5%	70%
56	Administrative & Waste Services	297	14%	30%	20%	35%
61	Educational Services	27	0%	0%	22%	78%
62	Health Care & Social Assistance	301	0%	10%	21%	70%
71	Arts, Entertainment & Recreation	97	13%	8%	61%	18%
72	Accommodation & Food Services	521	36%	20%	23%	21%
81	Other Services (except Public Administration)	293	1%	8%	22%	70%

The number of specialized places are based on aggregating 4 digit results to 2 digit industry sectors and percentages in columns are based on the number of specialized places.

Table 6: Specialization by Urbanization for Specific Industries

NAICS	Industry Name	Sets of Specialized Places	Min	Max	Mean	Std Dev
<i>Industries in Least Dense Specialized Places</i>						
5619	Other Support Services	1	11	117	61.9	34.6
6111	Elementary & Secondary Schools	2	278	297	287.5	13.4
7212	Recreational Vehicle Parks & Recreational Camps	2	18	1739	760.9	431.5
4442	Lawn & Garden Equipment & Supplies Stores	1	510	1203	821.6	215.5
3253	Agricultural Chemical Manufacturing	1	235	1790	880.3	404.4
7114	Agents & Managers for Public Figures	1	474	1655	918.6	350.7
5174	Satellite Telecommunications Services	1	74	2017	962.2	476.6
3159	Apparel Manufacturing	1	421	1596	1020.5	344.2
3365	Railroad Rolling Stock Manufacturing	1	737	1542	1054.7	182.9
5621	Waste Collection	4	235	2420	1097.1	673.1
<i>Industries in Most Dense Specialized Places</i>						
5151	Radio & Television Broadcasting	1	2597	2601	2599.2	1.7
7111	Performing Arts Companies	1	2590	2601	2595.2	4.6
5259	Other Investment Pools & Funds	1	2582	2601	2594.2	6.8
5418	Advertising & Related Services	1	2582	2601	2592.8	6.1
5414	Professional, Scientific & Technical Services	1	2581	2601	2592.5	6.2
8133	Social Advocacy Organizations	1	2561	2601	2590.8	10.8
4421	Furniture Stores	1	2590	2590	2590.0	0.0
6243	Vocational Rehabilitation Services	1	2561	2601	2588.9	10.1
8139	Business, Professional & Labor Organizations	1	2560	2601	2588.2	12.0
8132	Grantmaking & Giving Services	1	2550	2601	2585.0	13.7
<i>Industries with Highest Variance in Density of Specialized Places</i>						
5511	Management of Companies & Enterprises	4	549	2600	1984.9	846.5
5629	Remediation & Other Waste Management Services	2	507	2508	1654.4	732.5
3161	Leather & Hiding Tanning & Finishing	1	22	2238	1142.9	717.4
4812	Nonscheduled Air Transportation	2	541	2566	1665.1	699.3
5161	Internet Publishing & Broadcasting	3	543	2449	1714.1	696.4
8129	Other Personal Services	3	882	2601	1878.2	696.0
5621	Waste Collection	4	235	2420	1097.1	673.1
5324	Commercial, Industrial Machinery Rental	2	854	2546	1779.8	652.7
4842	Specialized Freight Trucking	4	19	2508	1317.3	614.7
4881	Support Activities for Air Transportation	3	219	2539	1362.4	611.2

Values for Min, Max, Mean and Std Dev are based population density of all specialized places in a given industry. The most dense place is given a rank of 2,601 and the lease dense place a rank of 1.

Table 7: Specialization and Localization

NAICS Code	Industry Name	No. 4 digit Industries	% of 4 digit Industries Subject to Specialization	% of 4 digit Industries Localized
31-33	Manufacturing	78	41.0%	15.4%
42	Wholesale Trade	19	94.7%	52.6%
44-45	Retail Trade	27	74.1%	25.9%
48-49	Transportation and Warehousing	23	69.6%	39.1%
51	Information	16	56.3%	37.5%
52	Finance & Insurance	11	72.7%	54.5%
53	Real Estate & Rental & Leasing	8	75.0%	25.0%
54	Professional, Scientific & Technical Services	9	88.9%	44.4%
55	Management of Companies & Enterprises	1	100.0%	100.0%
56	Administrative, Waste & Remediation Services	11	63.6%	18.2%
61	Educational Services	7	42.9%	14.3%
62	Health Care & Social Assistance	18	66.7%	27.8%
71	Arts, Entertainment & Recreation	9	55.6%	22.2%
72	Accommodation & Food Services	7	85.7%	14.3%
81	Other Services (except Public Administration)	14	64.3%	42.9%
	Total	258	62.0%	28.7%

Percent Localized is based on the methodology described in Section 8.1 and described in more detail by Duranton and Overman (2005)

Table 8: Localization by Sets of Specialized Places

# Sets of Specialized Places	Non-Localized Industries subject to Specialization		Localized Industries subject to Specialization	
	4-digit Industries	% Establishments	4-digit Industries	% Establishments
1	50	15.7%	47	26.7%
2	22	16.7%	15	36.4%
3	9	15.3%	7	39.7%
4	5	18.1%	2	27.6%
5	<u>1</u>	55.0%	<u>1</u>	39.0%
All	87	16.4%	72	30.6%

Results just include those industries subject to specialization and % Establishments indicates the portion of all establishments in specialized places.

Figure 1: Population Kernel - All Industries

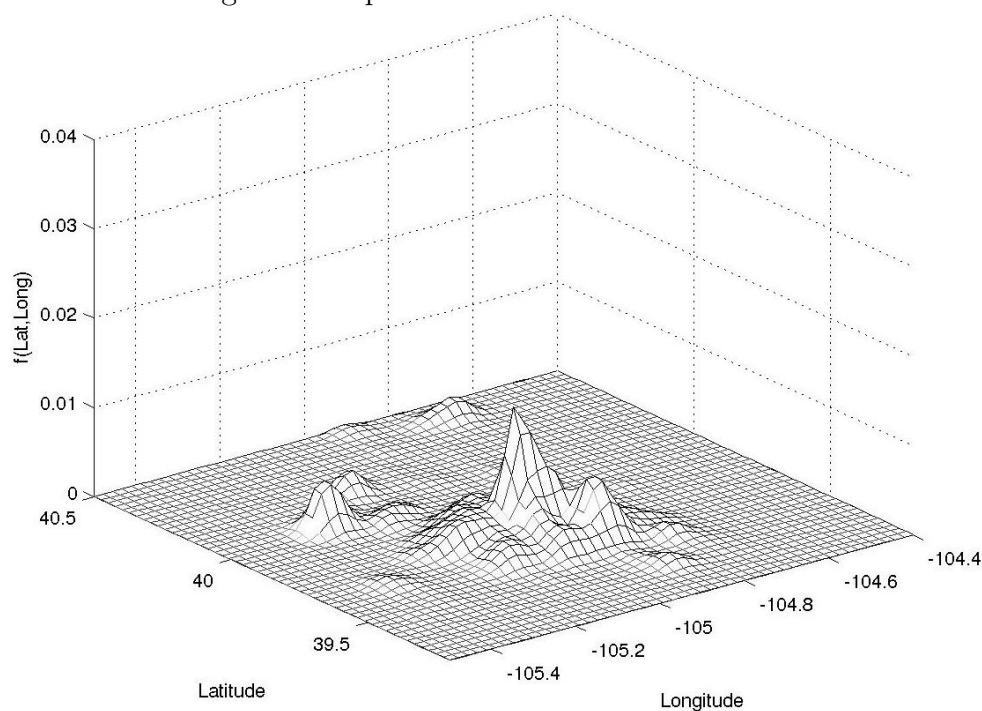
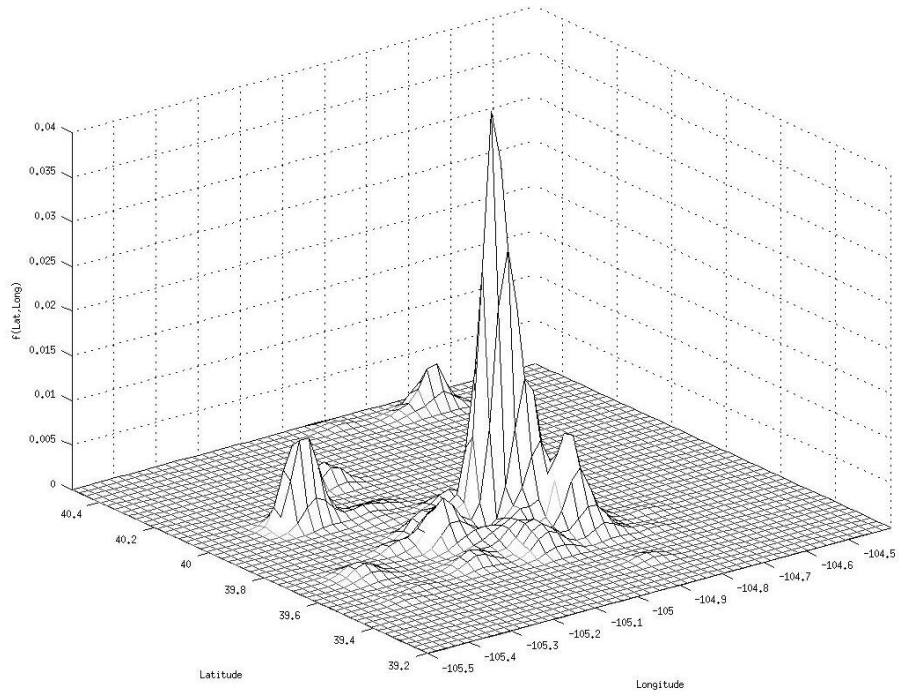
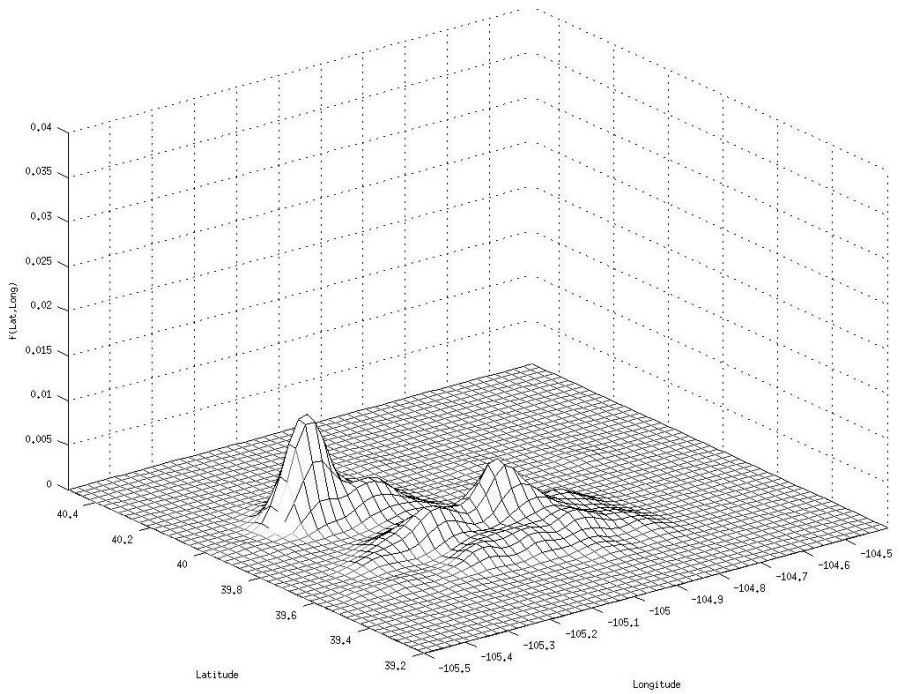


Figure 2: Population Kernels

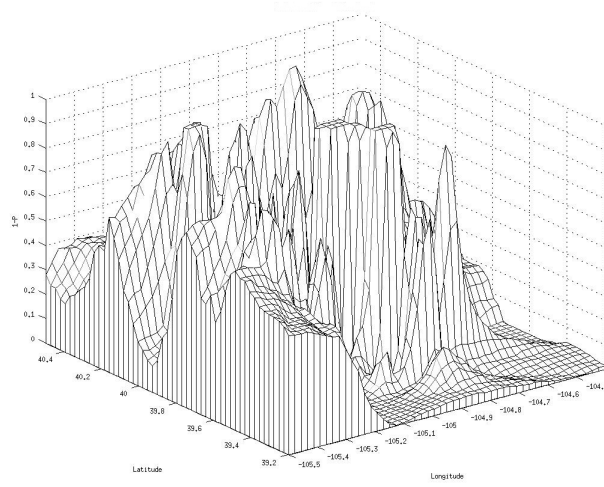


(a) NAICS 5411 Legal Services

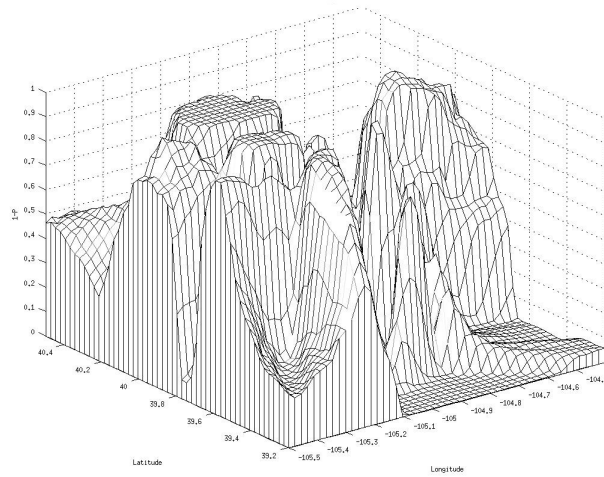


(b) NAICS 5417 Scientific Research and Development Services

Figure 3: Local P-values ($1 - p_i^{local}$)

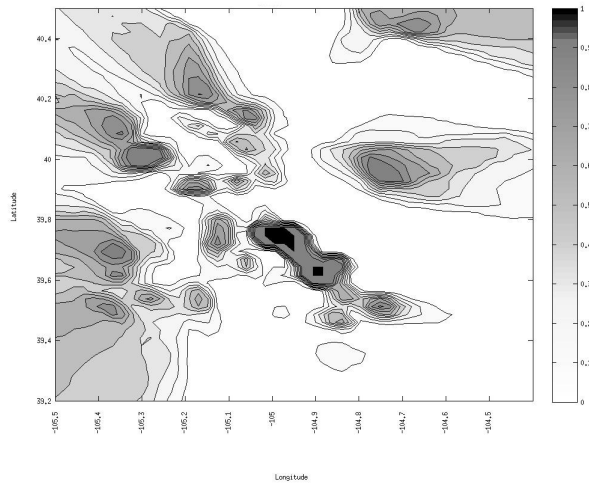


(a) NAICS 5411 Legal Services

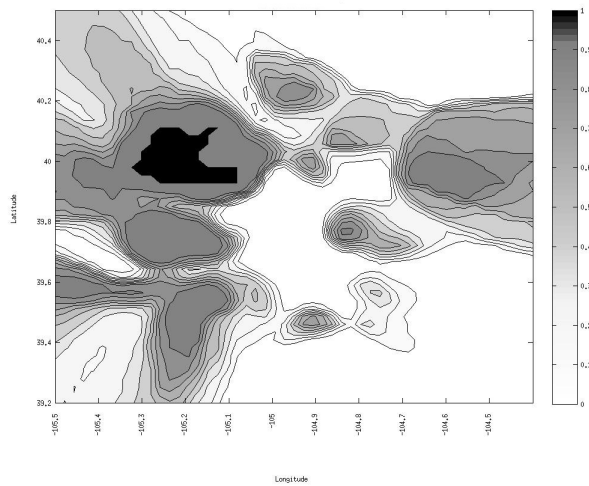


(b) NAICS 5417 Scientific Research and Development Services

Figure 4: Local P-value Contours

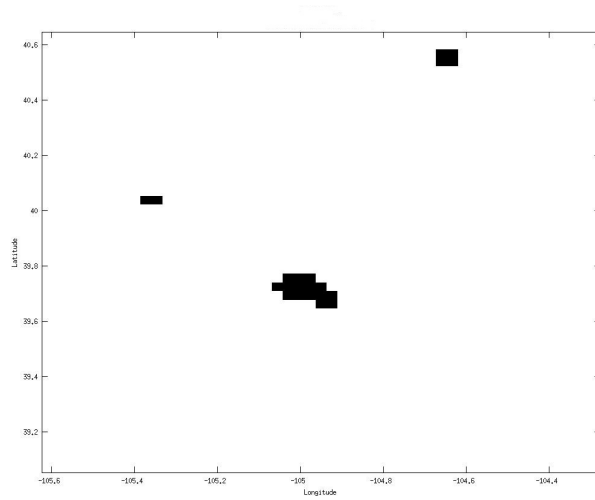


(a) NAICS 5411 Legal Services

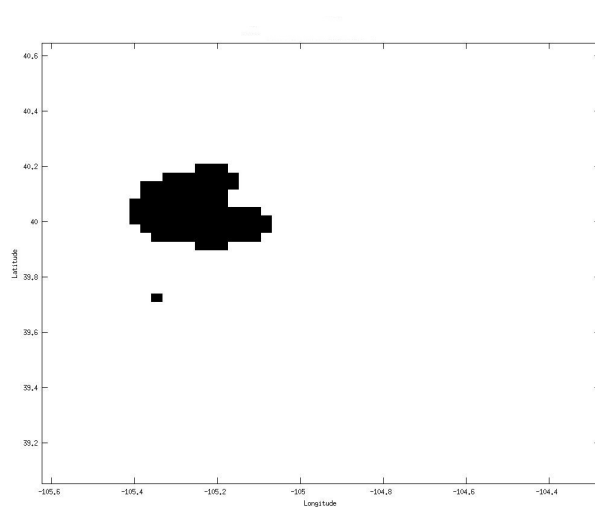


(b) NAICS 5417 Scientific Research and Development Services

Figure 5: Globally Significant Specialized Places



(a) NAICS 5411 Legal Services



(b) NAICS 5417 Scientific Research and Development Services

Figure 6: Global P-Values

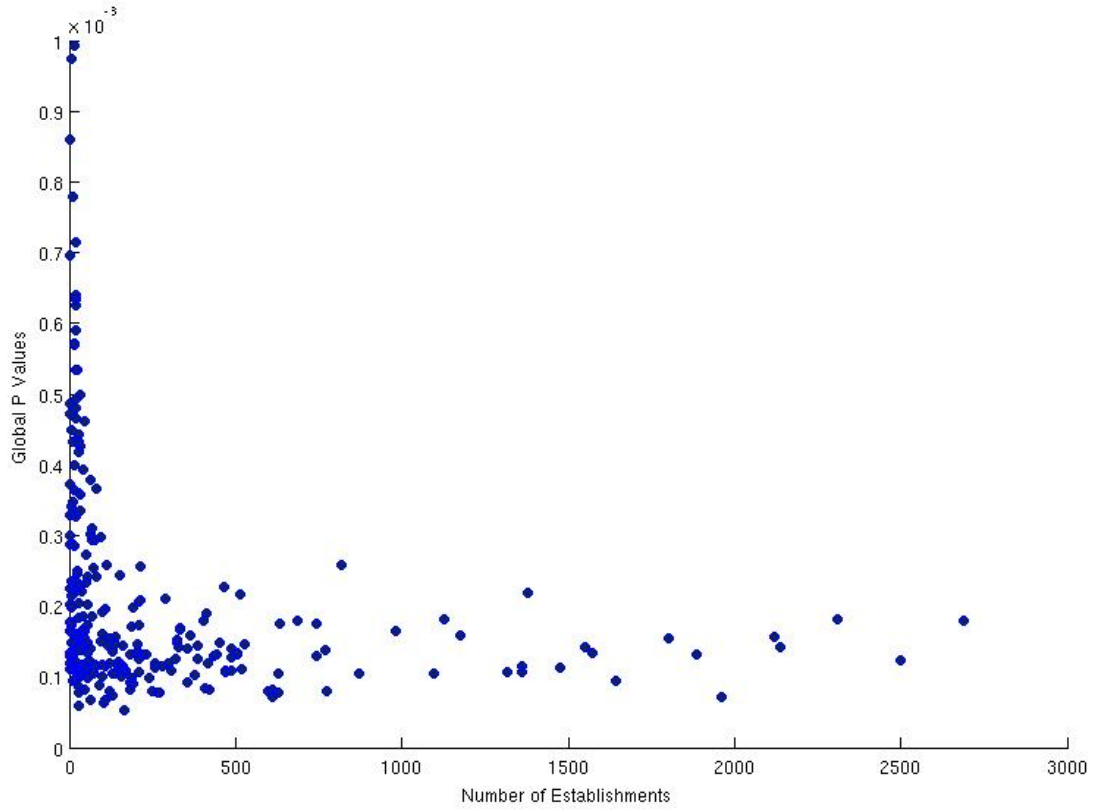


Figure 7: Total Number of Industries Subject to Specialization by Place

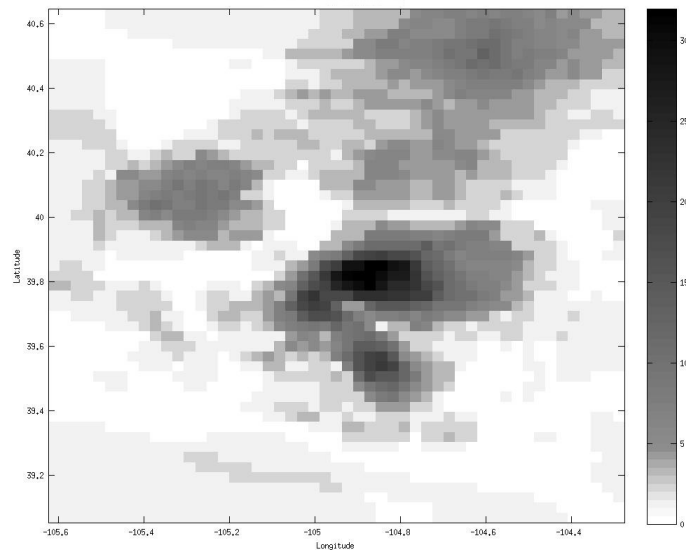
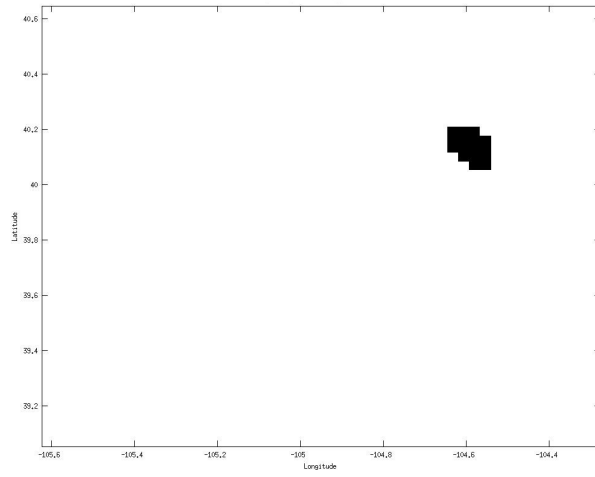
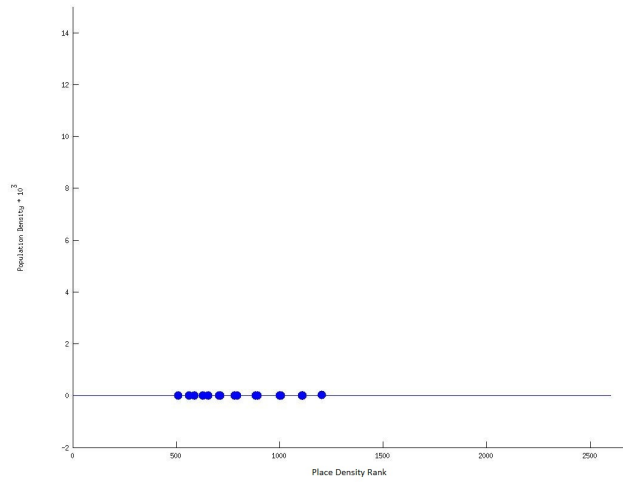


Figure 8: Lawn & Garden Equipment & Supplies Stores

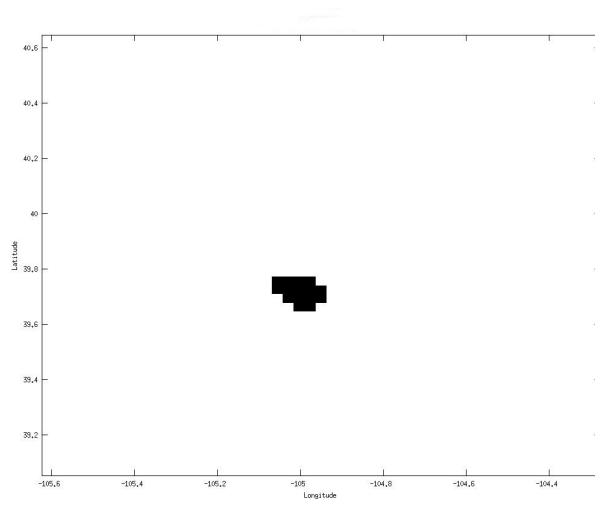


(a) Specialized Places

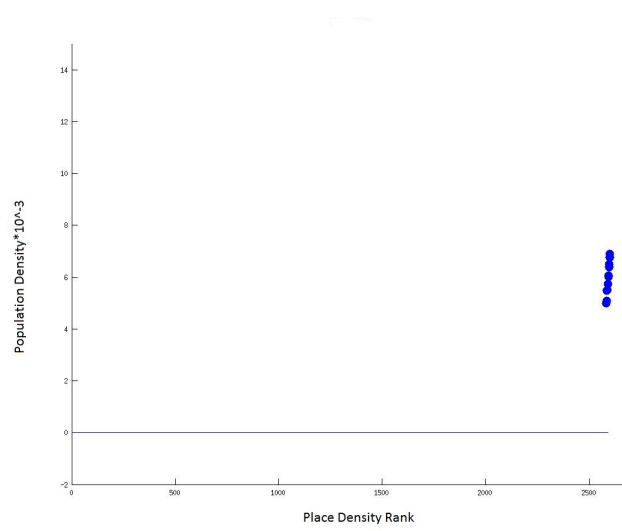


(b) Specialized Places Intensity Rank

Figure 9: Advertising and Related Services

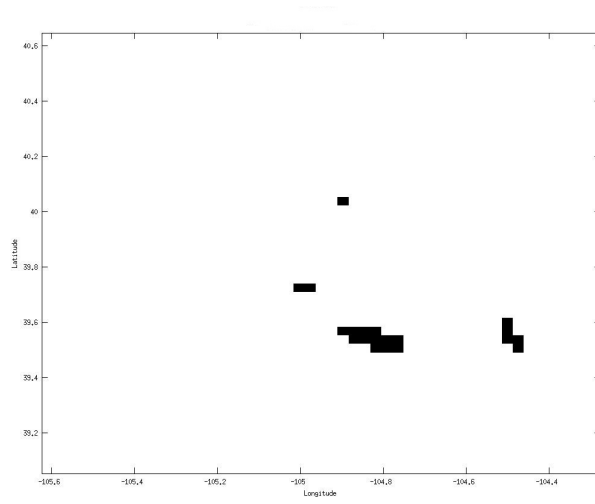


(a) Specialized Places

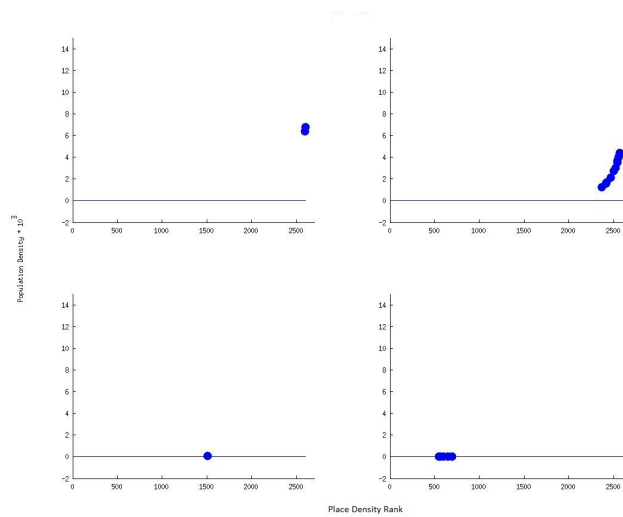


(b) Specialized Places Intensity Rank

Figure 10: Management of Companies & Enterprises

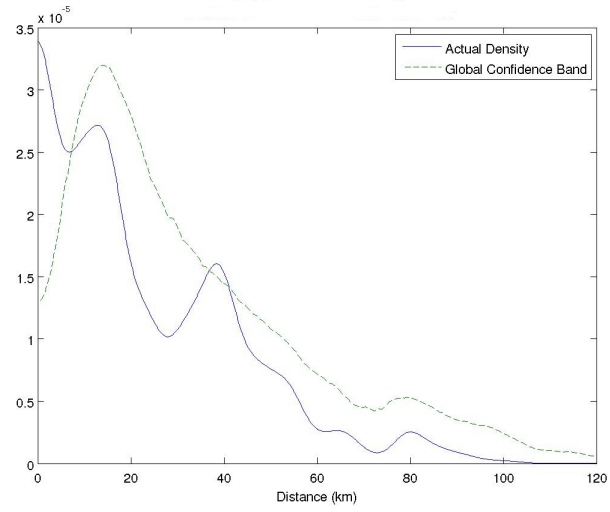


(a) Specialized Places

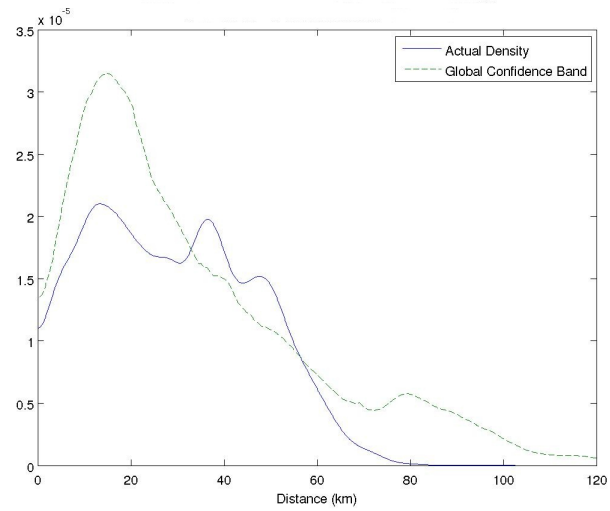


(b) Specialized Places Intensity Rank

Figure 11: Duranton & Overman Test for Localization

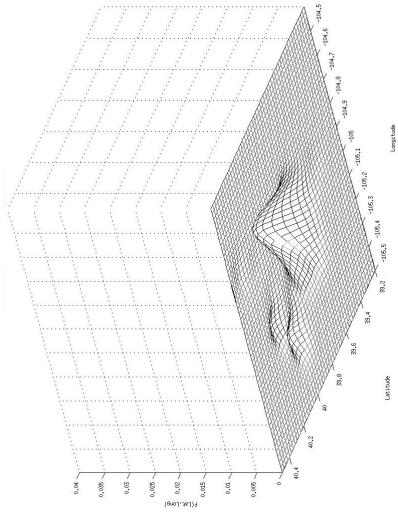


(a) NAICS 5411 Legal Services

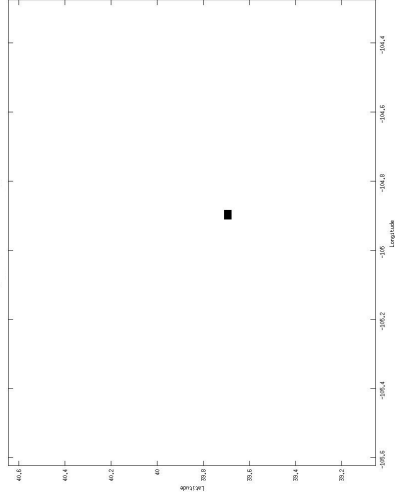


(b) NAICS 5417 Scientific Research and Development Services

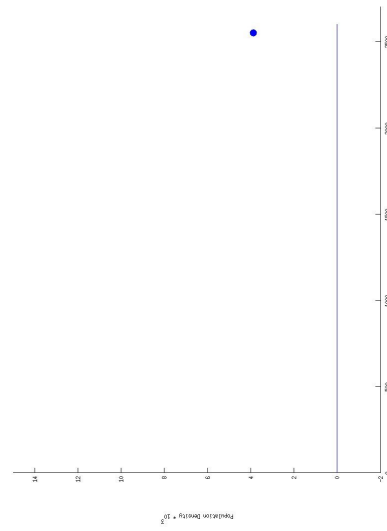
Figure 12: NAICS 3118 Bakeries & Tortilla Manufacturing



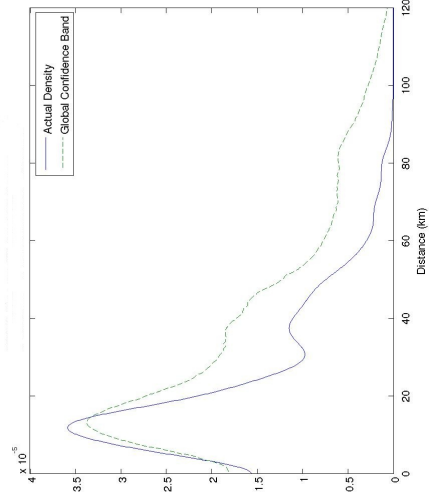
(a) Industry Kernel



(b) Globally Significant Specialized Places

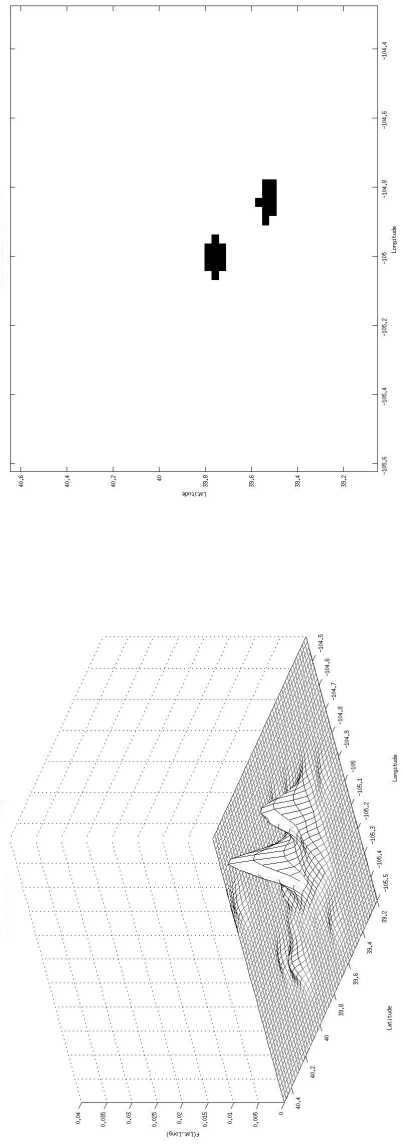


(c) Specialized Places Plots

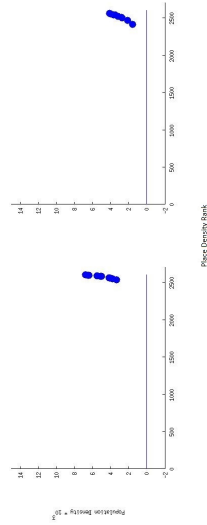


(d) Test for Localization

Figure 13: NAICS 5171 Wired Telecommunications Carriers

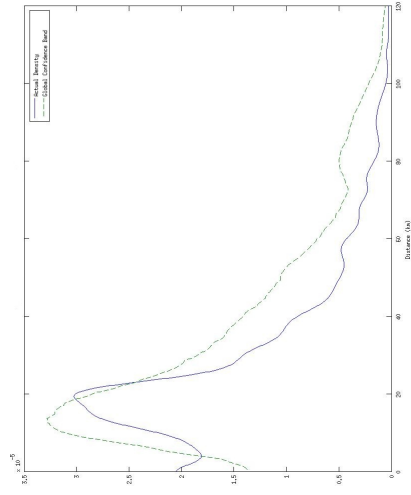


(a) Industry Kernel



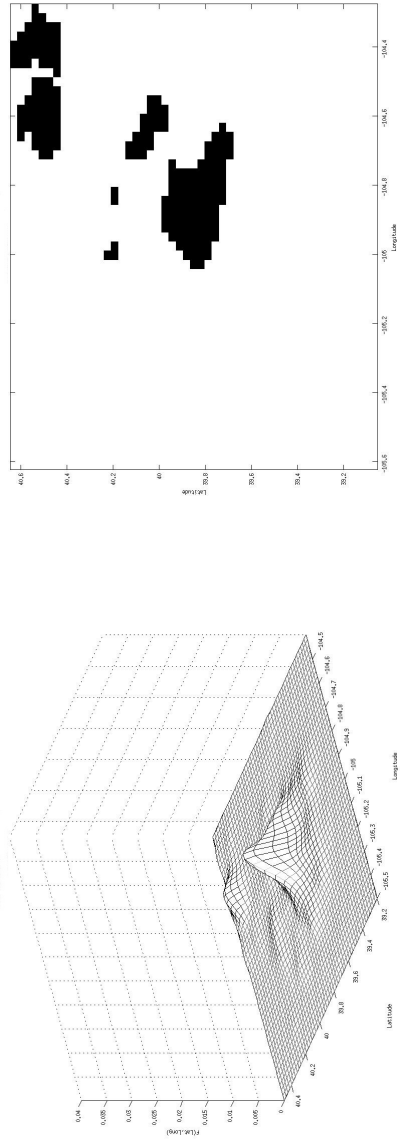
(c) Specialized Places Plots

(b) Globally Significant Specialized Places



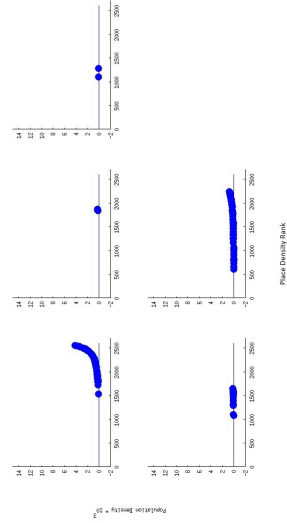
(d) Test for Localization

Figure 14: NAICS 4841 General Freight Trucking

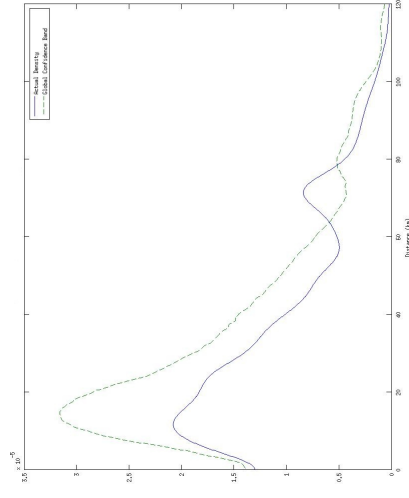


(a) Industry Kernel

(b) Globally Significant Specialized Places

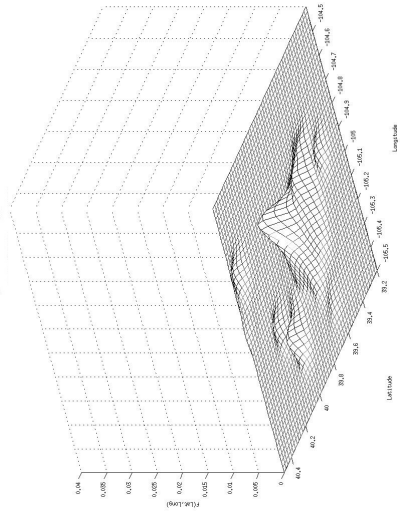


(c) Specialized Places Plots

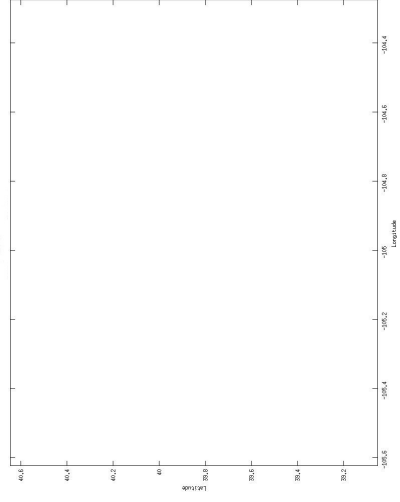


(d) Test for Localization

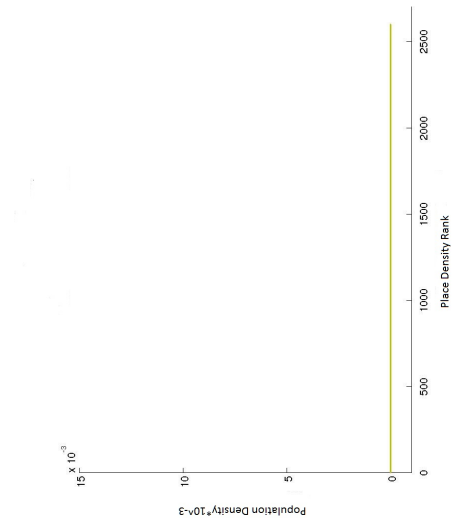
Figure 15: NAICS 4422 Home Furnishing Stores



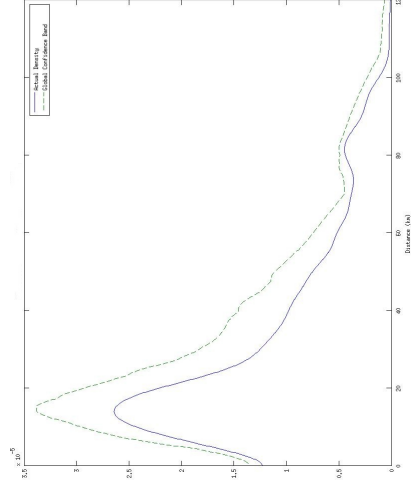
(a) Industry Kernel



(b) Globally Significant Specialized Places



(c) Specialized Places Plots



(d) Test for Localization