DRAFT ONLY -- Do Not Cite Without Permission of the Authors

The Topography of Metropolitan Employment:
Identifying Centers of Employment in a Polycentric Urban Area*

by

Christian L. Redfearn

School of Policy, Planning, and Development
University of Southern California

March 6, 2005

Abstract

Increasingly, U.S. metropolitan areas are polycentric. While this is well recognized, there is little consensus as to the appropriate method for identifying centers of employment and their extent. Discussions of sprawl and decentralization, agglomeration and productivity, and the impacts of transportation or land-use regulation on urban structure depend crucially on the spatial accounting of employment within a metropolitan area. Existing methods for subcenter identification suffer from strong assumptions about parametric form, misspecification, or reliance on local knowledge to calibrate model parameters. Using data from the greater Los Angeles metropolitan area, this paper introduces a nonparametric method for identifying subcenters – both their centroids and boundaries. This method is benchmarked against representative alternatives for subcenter identification. The importance of the difference in approaches is made clear by comparing their measured concentration of the greater Los Angeles metropolitan area. Results indicate that this, more flexible, nonparametric approach yields both greater accuracy in defining subcenter boundaries and better resolution identifying a wide range of subcenters. These attributes should better inform research that employs density as an independent or dependent variable.

Key Words: Urban Economics, Spatial Econometrics, Subcenters

*This paper has benefited from the advice of Genevieve Guiliano, Raphael Bostic, and the participants of the Lusk Center Real Estate Seminar series on October 29, 2004.

1 Introduction

The starting point for the empirics of urban structure is the monocentric model. The simplicity of von Thünen's "isolated state" is the basis for an elegant and historically successful model of the distribution of residential location around a single node of employment at the center of a city.[1] A single-centered model may still be appropriate for smaller cities, but most larger urban areas are now polycentric (McMillen and Smith 2003). Accordingly, multi-centered models are necessary to capture accurately the system of subcenters within this growing number of metropolitan areas. Current methods for identifying subcenters fall into three broad categories: clustering models, parametric models of increasing complexity, and nonparametric models. While clustering methods rely on *ad hoc* definitions for density cutoffs and parametric models make strong assumptions regarding physical symmetry, collectively they have demonstrated convincingly the polycentric nature of metropolitan areas and the importance of explicitly accounting for the system of subcenters within them. The model introduced in this paper is nonparametric – incorporating a fully flexible form in order to capture the irregular distribution of urban employment within an urban area.

Accurately defining subcenters is important because many of the most interesting questions in urban economics employ urban structure either as an independent or dependent variable. For example, employment density may be used to explain productivity, while residential density may be correlated with the incidence of crime. Alternatively, transportation costs or land-use regulation can be used to explain either employment or residential density. Because the underlying location of activity within many metropolitan areas has evolved to concentration in two or more subcenters, models that fail to incorporate this polycentric reality will inform these questions poorly.

[1] Monocentric models have also been used to describe firm location and are still widely used in residential applications. In fact, Bertaud and Malpezzi (2003) argue that the negative exponential parameterization of the monocentric model can hold so long as there is "a higher density of employment at the center than elsewhere." (page 5).

On the question of sprawl, for example, the measurement of concentration has frequently been measured using simple parametric models that employ a single density gradient to assess trends in dispersion. A flattening of the density gradient over time is taken as evidence of sprawl when numerous scenarios of concentration are consistent with the observation. An alternative explanation for density gradients becoming less steep is greater concentration within subcenters outside the CBD. While the center may be losing relative importance, it could be that aggregate density and/or concentration is greater. An inference of dispersal based on this approach would be premature and stems from model misspecification – a monocentric model in a polycentric context.

The move to polycentric models is important both because it may be more appropriate and because doing so offers additional sources of variation which may be exploited in addressing other empirical questions. In some applications, it may be reasonable to use the city as the unit of observation. But for many empirical questions, this approach collapses a system of intra-urban subcenters into a single point, potentially wasting variation across subcenters. Giuliano and Small (1991) find 32 subcenters in the five-county Los Angeles consolidated metropolitan statistical area (CMSA) – do they have common employment composition? Are they equally productive? Are they growing at the same rate? Subcenters offer a rich source of variation all while holding constant weather, state laws, and other causes of concern that accompany variation found by looking across metropolitan areas.

In order to exploit intra-urban variation in density – either residential or employment – a convincing procedure for identification of subcenters and their extent is required. The goal of this research is to develop such a method that both overcomes the strong assumptions of symmetry embedded in many of the parametric models and is sufficiently general to obviate the need for calibration on a city-by-city basis.

The method developed in this paper is straightforward and can be described with surprising ease, beginning with the estimation of nonparametric employment density surface.

The window size is kept sufficiently small to keep intact the structure of local employment density necessary to identify subcenters. Local maxima on the density surface become candidate centers. Their boundaries are defined along rays from the local maxima to a point where the surface becomes flat or reaches an inflection point, which indicates an adjacent subcenter. The set of candidates is then tested to determine whether proximity to the center is statistically significant using a negative exponential model on tracts within and just adjacent to the subcenter. Each subcenter is defined by the tracts within this flexibly-determined boundary. This flexibility is an important strength relative to parametric approaches. In practice, subcenters follow transit corridors, natural features, or abut other subcenters – they rarely are concentric circles or other shapes that are readily parameterized.

The data are drawn from the Los Angeles CMSA, which proves an excellent testing ground; the greater metropolitan area offers both many subcenters and many of unusual shape. The nonparametric approach finds 31 statistically significant employment subcenters in the Los Angeles CMSA. The results also indicate that the nonparametric procedure identifies small centers – centers with density and total employment below the cutoffs used in the clustering approaches to subcenter identification.

The nonparametric procedure yields the centroid and boundary of each center within a system of subcenters. Such a system offers the opportunity to study the origin and dynamics of urban concentration and its role in a host of urban phenomena. For example, the complete mapping of the internal distribution of employment could be used to examine the importance of centrality; the improved definition of subcenters allows for a more accurate accounting of the types of employment within centers of different sizes – speaking to the nature of agglomerative economies; and, in a panel context, the system of subcenters could be tracked over time to understand better the internal dynamics of urban evolution. Moreover, nothing in the procedure is employment-specific. While employment is the focus of the applications in this paper, it could equally well be applied to residential density or other spatially-varying phenomena.

The paper is organized as follows: Section 2 introduces polycentricity in Los Angeles and describes the data used in this analysis. Section 3 formally develops the nonparametric method introduced above. Other comparison approaches are discussed in Section 3. Contrasts between the various approaches are drawn in Section 4. Section 5 concludes and discusses several extensions.

2 Employment in Los Angeles

Los Angeles is the setting for the empirical testing of the nonparametric method introduced in this paper. Relatively late-developing and organized around the automobile, it is the proto-typical polycentric metropolitan area. Its network of internal freeways, multitude of jurisdictions,[2] and multi-racial/multi-ethnic nature (there is no majority racial or ethic group in Los Angeles as of 2000) provide ample opportunities for urban theorists and urban empiricists alike. Many of the issues facing metropolitan areas across the country were first observed in Los Angeles, recently noted examples include sprawl, ethnic/racial plurality, and polycentricity.[3]

The data used in this examination of polycentricity were provided by the Southern California Association of Governments (SCAG) and consist of Census tract-level employment and population data for the 2000 urbanized area portion of the five county Los Angeles Consolidated Metropolitan Statistical Area (CMSA). These employment data are developed by SCAG from wage and compensation survey data reported to the State Economic Development Department (EDD) of the California Labor and Workforce Development Agency. The EDD data includes all jobs subject to wage and tax reporting, but excludes self-employment

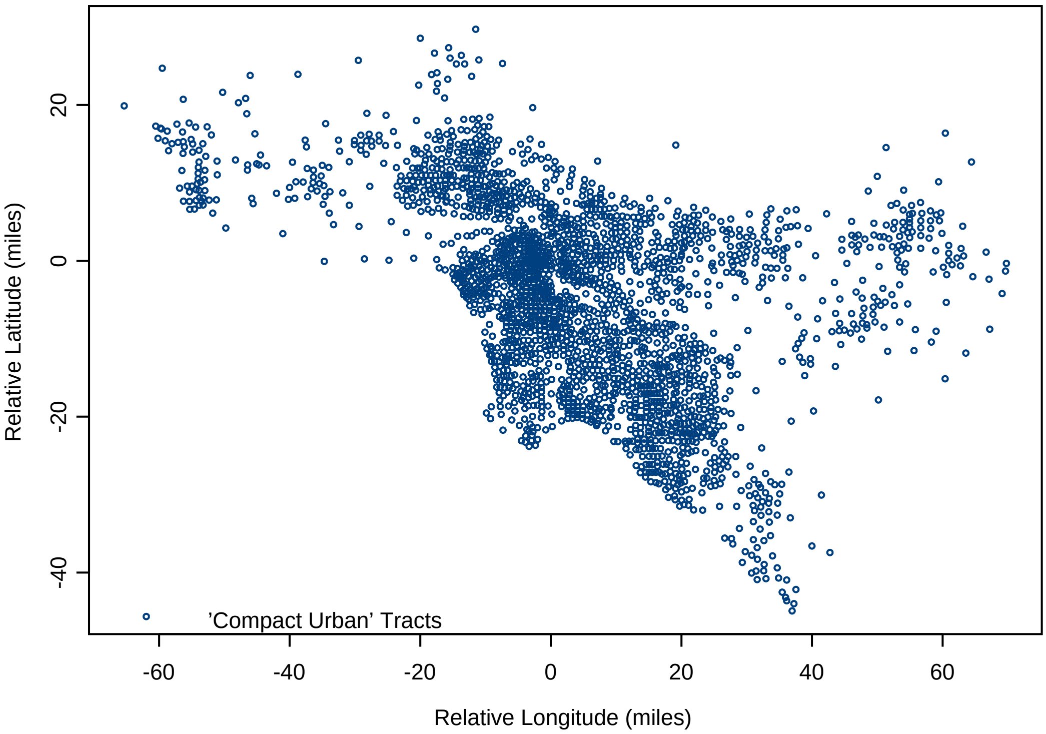
[2]There are 166 member cities in the six-county Southern California Association of Governments, these counties are the five in the Los Angeles CMSA and San Diego.

[3]In fact, Gordon and Richardson (1996) argue that Los Angeles is "past polycentricity," and has moved on to become a "dispersed metropolis."

and private household workers.[4]

The Los Angeles CMSA is comprised of five counties: Los Angeles, Orange, Riverside, San Bernardino and Ventura. Though defined by the U.S. Census as urbanized,[5] a small number of tracts at the edges of the urbanized area were too large and too empty to be useful in the nonparametric regressions. Because of this, the set of urbanized tracts was further limited by requiring tracts to be proximal and sufficiently dense. The resulting set of 2,410 "compact urban" tracts was used for all of the analysis presented in this paper. Figure 1 plots the tracts. It is clear that the urban tracts are sufficiently small in area to easily trace

Figure 1: Location of 'Compact Urban' Tracts



the coastline of the Pacific Ocean to the south and west. The suburban tracts provide less

[4]See `http://www.calmis.ca.gov/file/resource/indmeth.htm` for more information. The exclusion of self-employment and household workers may represent a large number in aggregate. However, it is not clear that it is distributed such that it would bias the estimated system of subcenters. Moreover, any bias would not influence comparisons between methods for identifying subcenters – all use the same data.

[5]See `http://www.census.gov/population/censusdata/urdef.txt` for more information.

precision in defining the northern and eastern boundaries created by mountains and desert. The total area covered by the tracts is just over 3.2 million acres (5,000 square miles), and includes approximately 5.4 million jobs.

Table 1 gives an indication of the degree of concentration in the CMSA and across its

Table 1: Tract Area-Employment Distributions by County

	Cumulative % Total Employment					
Cumulative % Total Area	Los Angeles CMSA	Los Angeles County	Orange County	Riverside County	San Bernardino County	Ventura County
10	0.0	0.0	0.2	0.5	0.3	0.0
25	0.1	0.1	1.3	0.5	0.7	0.0
50	0.6	0.7	8.0	3.0	7.0	0.4
75	7.2	6.1	25.6	22.8	31.3	15.5
90	29.4	26.3	48.1	59.7	60.9	42.8
95	47.9	45.2	63.2	71.1	73.2	59.5
98	67.8	66.6	74.9	83.6	86.5	77.9
99	77.7	78.4	78.7	95.9	93.0	84.5
100	100.0	100.0	100.0	100.0	100.0	100.0
Obs.	2405	1609	478	62	130	126

five constituent counties. The table reports the share of total employment by share of total land when tracts are ordered by increasing density. For example, if the total land area in the compact set of tracts used in this analysis were split 75%/25%, with no tract in the first portion more dense than any tract in the second, the first portion would contain only 7.2% of total employment. This is seen in the table by reading down the first column to the row reporting the results for the 75th percentile of total area.

The degree of concentration reported in Table 1 is striking. In the five-county region, 22.3% of all employment is crowded onto 1% of the total sample tract area. Not surprisingly Los Angeles county is the most dense. The Orange County statistics reflect its transition from suburban to urban. The remaining three counties are largely suburban and rural and have a more evenly spread employment distribution.

What cannot be shown in this table is the spatial distribution of density. This is important for two reasons. First, Table 1 can't show the clustering of the densest tracts; the extensive density in Los Angeles county may be concentrated in one subcenter or each of these dense tracts may itself be its own center. There are just over 300 tracts with density greater than 45 jobs per, how they are organized spatially is not apparent. Second, because subcenters are defined by their employment density relative to their surroundings, these 300 highly-dense tracts say nothing about the distinct subcenters outside the core of downtown Los Angeles. It is essential therefore to know the spatial distribution of all employment, not just the that found in the densest part of the urban area.

3 The Topography of Urban Employment

That metropolitan areas are polycentric is not in question. There is ample empirical evidence that polycentrism is the rule in larger cities, not the exception (Giuliano and Small 1991, Hughes 1993, Cervero and Wu 1997, McMillen and McDonald 1998).[6] At issue is the identification of subcenters within the metropolitan area – including both their location (centroid latitude and longitude) and extent (their boundaries). This section develops a flexible nonparametric approach to subcenter identification and discusses its strengths and weaknesses *vis á vis* several existing methods.

The goal is to locate and describe concentrations of employment that are significantly more dense than employment in the surrounding areas. While diminishing somewhat in relative importance, downtown remains the dominant subcenter within a metropolitan area (Anas, Arnott, and Small 1998). Downtowns are easy to identify and readily fit the criteria of significantly higher density than their surroundings. A successful procedure will identify and bound not only the city center but the entire hierarchy of subcenters within a metropolitan

[6] See Anas, Arnott, and Small (1998) for an excellent review of empirical evidence and theoretical research on subcenters and subcenter formation.

area.

3.1 A Nonparametric Model of Subcenter Identification

Empirical research on modeling urban structure has moved towards greater flexibility to move closer the irregular reality of cities. The logical next step is full flexibility. The nonparametric model proposed in this paper requires neither local knowledge to calibrate density cutoffs nor strong assumptions about the spatial symmetry of subcenters. The cost of these advantages comes in terms of simplicity: whereas the mono- and polycentric models offer a simple metric of centrality – the density gradient – there is no equivalent metric for the nonparametric model. But while some additional calculation may be necessary, the payoff is the identification of the system of subcenters with boundaries that much more closely reflect their true structure. This should result in more appropriate clustering of tracts than under the alternative methods.

The first step in the process is to estimate a density surface using locally-weighted regression (loess). Locally-weighted regression involves the estimation of two slope parameters at each point on a grid overlaying the sample area – in this case, the set of "compact urban" Census tracts in the Los Angeles CMSA. At each point on the grid, the two local regressions – one for the slope along the North-South axis, the other along the East-West axis – are estimated using a subset of all observations. Subsets are determined by the *span* – the fraction of all observations used in the local regressions (a span of 1 would use all the data). The span times the sample size yields the *window size* or number of observations used in the local regressions. Within the window, observations are weighted by a *kernel* function, with closer observations receiving more weight. The specific kernel used in this research is the commonly-used tri-cubic kernel function, given by:

$$(1) \quad weight = \left(1 - \left(\frac{dist_{ij}}{dist_i^{max}}\right)^3\right)^3 \quad ,$$

where $dist_{ij}$ is the distance from the point of the local regressions to tract j – a tract within in the window – and $dist_i^{max}$ is the largest distance from the point of interest to any point within the window. In this way, points receive decreasing weight – at an accelerating rate – as distance from the regression point increases; the furthest point receives no weight. The resulting coefficients at each grid point form a surface.[7]

The central tradeoff made when implementing the loess approach is between local accuracy and smoothness. The greater the window size, the smoother the resulting surface will be: as more observations are used in local regressions, the resulting surface will be less reflective of local structure because the additional observations will be drawn from further and further away. In this research, a span of .01 (one percent of the sample) is used to determine the subset of observations in each local regression. This amounts to employment density data from the 24 nearest tracts is used in each local regression. This is a small window size and leaves intact much of the local density structure required to identify candidate subcenters.

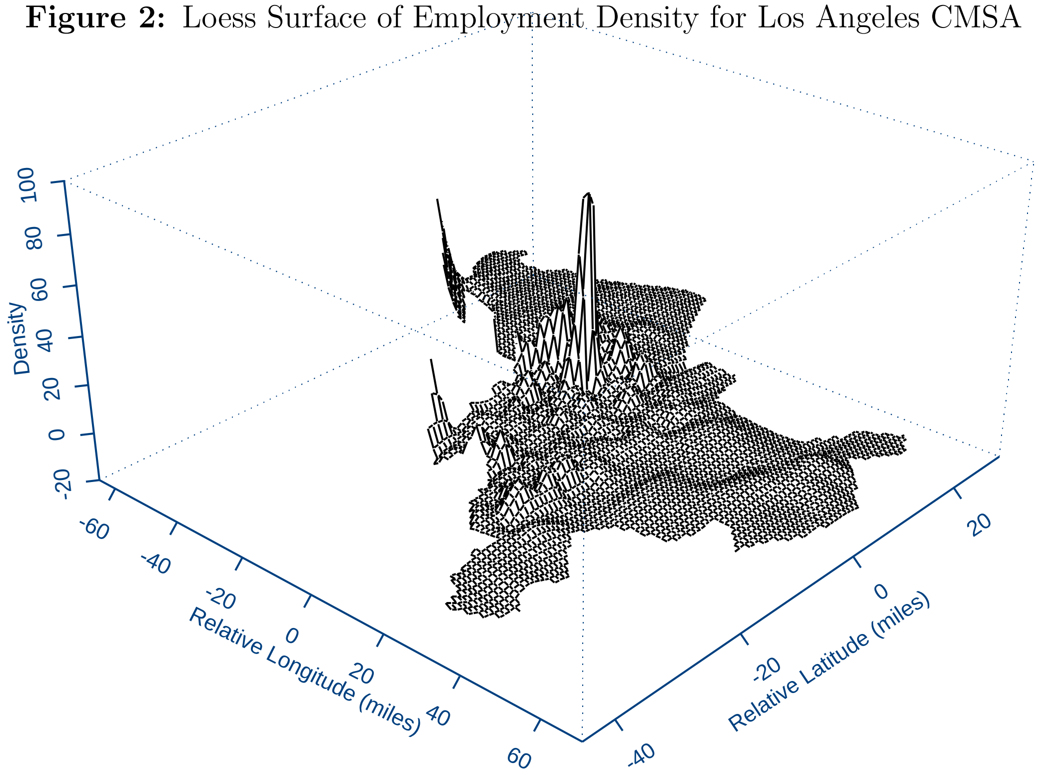
3.1.1 The Topography of Employment in Los Angeles

The result of these local regressions is the density surface shown in Figure 2. The most notable feature of the loess surface is the prominence of downtown. The consistent association in the popular media linking Los Angeles and sprawl obscures the fact that the downtown remains the largest employment center in the region and, by far, the most dense. However, it is also clear from the figure that there is a system of subcenters. The apparent density hierarchy moves from the highly densified downtown to a set of distinct peaks in its immediate vicinity. The direction of change in employment density however is not monotonically decreasing with distance from the CBD; other peaks and hills are scattered throughout the region.

Notable exceptions to a general decline in density away from the center are Ventura,

[7]For more on loess see Cleveland and Devlin (1988).

Figure 2: Loess Surface of Employment Density for Los Angeles CMSA



in the far northwest the beach cities southwest of downtown, and the string of Orange County cities, in the foreground of the figure. In the first two cases, the loess smoothing has not hidden the impact of a single very dense and influential tract. While these points on the surface appear odd – with estimated employment densities of 40 to 50 percent of downtown Los Angeles – they will not adversely impact the estimated location or extent of the respective subcenters: these tracts are, in fact, local centers. The robustness of the subcenter boundaries is discussed below. The string of cities that form the chain of peaks in the foreground are not due to idiosyncratic tracts. These concentrations of dense employment underscores the urbanizing nature of Orange County; these are the cities of Anaheim, Santa Ana, Costa Mesa, and Irvine.

Figure 2 begins to hint at one of the strengths of a nonparametric approach in this context. The system of subcenters is highly irregular in both their location and their shape.

That is, the surface indicates no "ring," "hub and spoke," or other simple classification for subcenter location. And, while harder to see, the figure also underscores the variability of subcenter shape: the traditional CBD of Los Angeles is a needle; the corridor of employment along Wilshire Boulevard appears as an east-west ridge; and, the two prominent rolls in the foreground follow freeways in the east of the region. No parametric approach will capture the richness of such an employment density surface.

3.1.2 Locating and Bounding Employment Subcenters

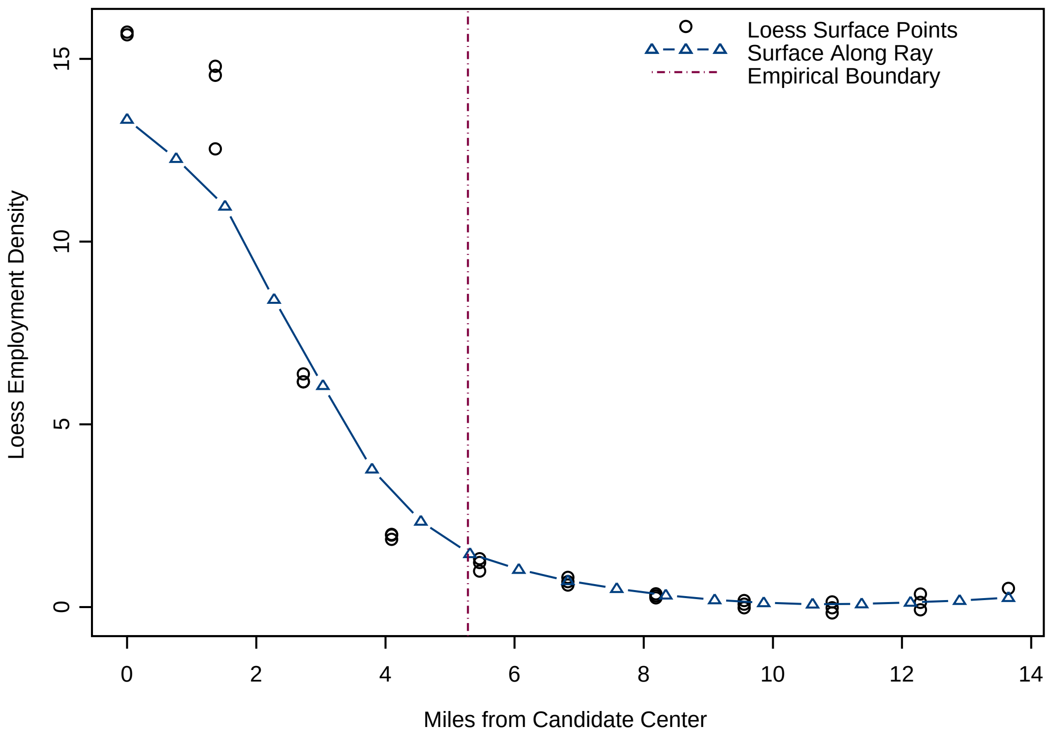
The next step is to identify candidate subcenters and define their boundaries. Locating candidates subcenters is straightforward: they are the set of local maxima on the loess surface. These are shown in Figure 2. There are 42 local maxima on the figure. Six of there are artifacts of of local regressions in areas sparsely populated with employment - small differences in local tract densities can lead to spurious peaks in the estimated surface. These six are eliminated due to too few proximal observations. That leaves 36 local maxima to be considered "candidate centers." Here local smoothing serves to clarify the local clusters of tracts with high employment density, but because of the small window size, individual subcenters are not lost to over-smoothing.

In keeping with the notion of flexibility, boundaries are defined by the slope of the loess surface along rays emanating from the centroid of each candidate subcenter. A circle around the candidate center is defined and partitioned into 32 rays which divide the surface into 11.25 degree wedges. The loess surface is defined at each of the points on the grid laid over the sample space, and so the surface along the rays that follow the directions of the main compass points and the angles that bisect them are well defined. However, the remainder of the rays are not. At each point on the ray, the nearest points are averaged to estimate the cross-sectional curve (amounting to a simple spatial moving-average).

Once defined, the rays then form the basis for determining the boundary of the subcenter.

For each of 32 rays, the slope is calculated iteratively until reaching an inflection point – indicating an adjacent subcenter – or until it becomes sufficiently flat to suggest that the subcenter no longer has an influence. Here flat is determined relative to the subcenter in question: flat is defined as one-half of the average slope between the local maximum and the minimum along the ray.[8] Figure 3 shows this process graphically. Points from the loess

Figure 3: Subcenter Boundary Determination



surface are represented by unconnected circles. The interpolated surface for this particular ray is illustrated by the sloped line/triangle combination. Along this ray, the boundary sits just beyond five miles from the candidate center. At this point the interpolated surface becomes sufficiently flat.

Figure 4 shows the same process for the Anaheim subcenter. The ray shown in the

[8]The total number of included tracts was robust to definitions of "flat" within a relatively large range around this one. Flat areas in the loess surface contain fewer tracts of larger area, moving the border of the subcenter often had no change on the underlying set of included tracts.

Figure 4: Subcenter Boundary Determination

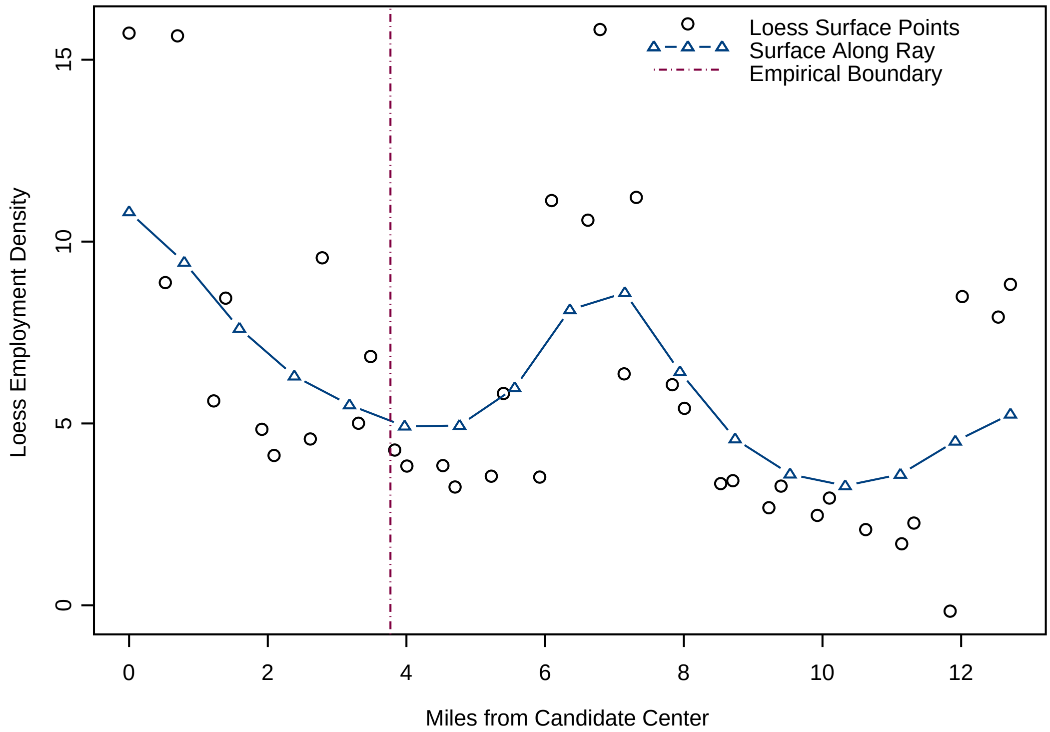
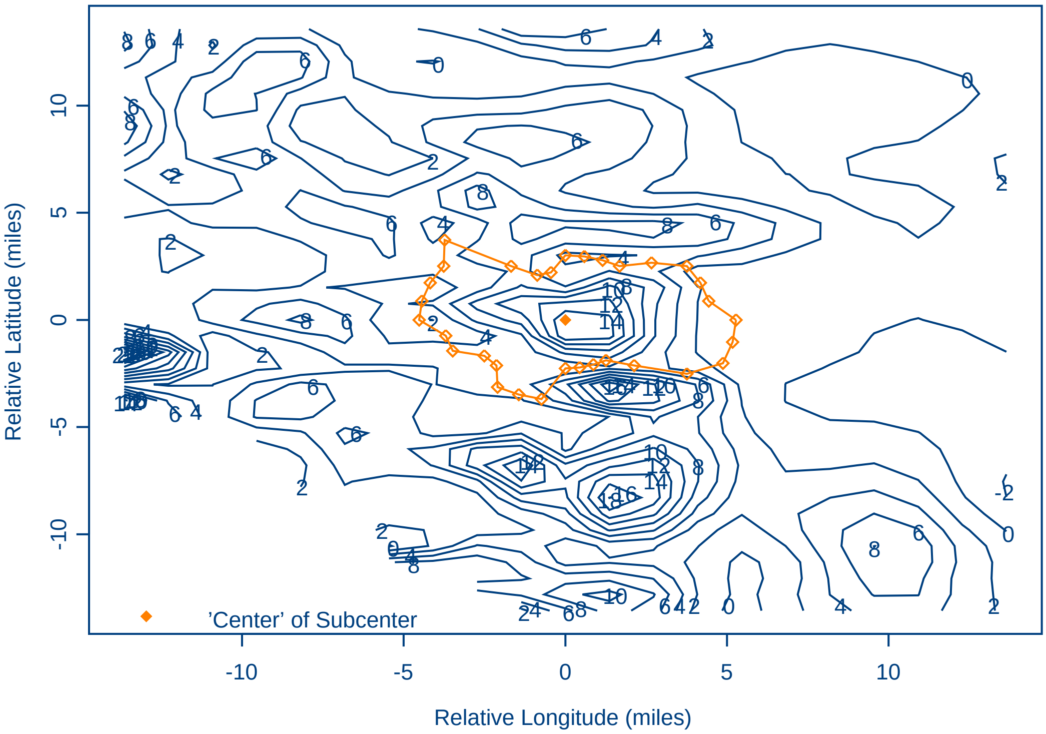


figure runs south towards Santa Ana, which is clearly visible as the density declines first moving away from the center of Anaheim before rising again when reaching Santa Ana. In this case, the boundary for Anaheim along this ray is based on an inflection point – on the change in slopes between adjacent subcenters. Note also the next in a string of employment concentrations can also be seen in Figure 4; the rise in density just beyond the twelve mile mark is Costa Mesa.

The definition of the border around the Anaheim subcenter is shown in Figure 5. The figure superimposes the 31 boundary points on a contour map of subcenter and immediate surrounding area. It is clear from the map that there are six distinct local maxima: Long Beach at the far west of the figure and the string of Santa Ana, Costa Mesa, Irvine and Newport Beach below Anaheim running north to south. It is also clear that none of the centers can be neatly described by concentric circles or other tidy parametric forms. Rather,

Figure 5: Boundary Definition for Anaheim Subcenter



the subcenters appear elongated, reflecting the fact that they are organized along significant freeways.[9]

The last step is to determine whether or not the candidate centers are in fact significantly more dense than their surroundings. To assess this, a simple negative exponential model is estimated using the tracts within the candidate subcenter and a set of adjacent tracts.[10] The basic criteria for adjacency is simple: tracts to be included in the regression are those within the subcenter boundary and those within a collar around the subcenter. Because the definition of a subcenter is a relative one – density should be significantly higher within a subcenter than outside it – it is potentially important to pick the surroundings data

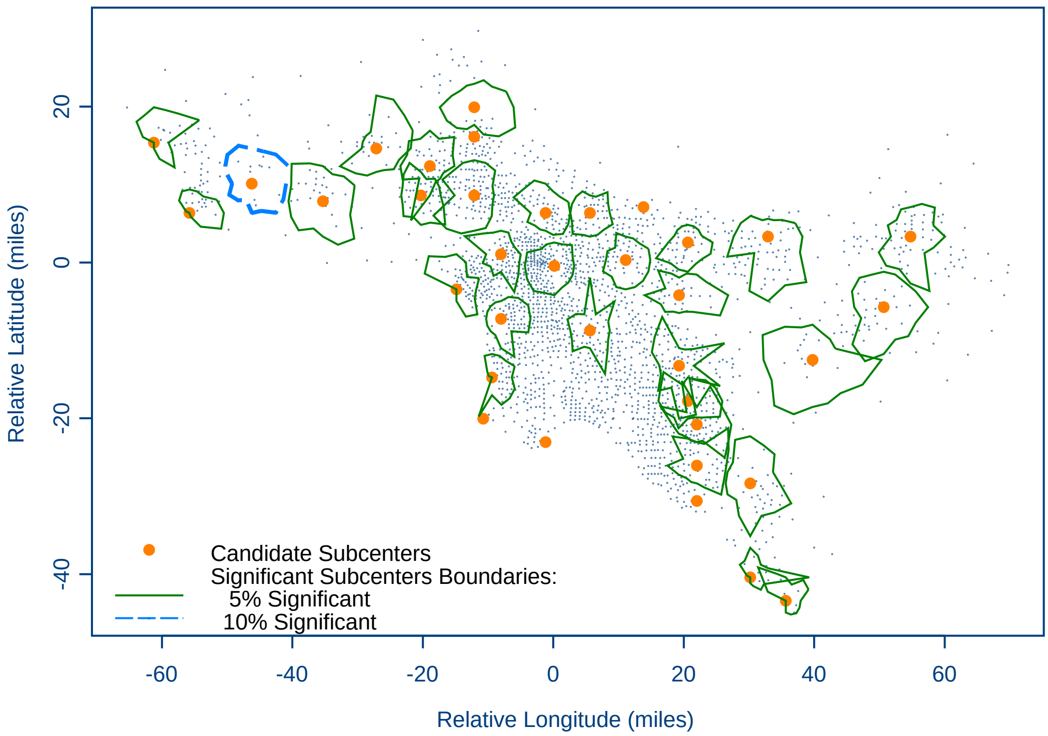
[9]These are Highway 91 through Anaheim, Highway 22 through Santa Ana, and Highway 405 through Costa Mesa and Irvine.

[10]This test works in favor of rejecting potential centers. All of the arguments in this paper for the use of a flexible approach suggest that use of a parametric model is inappropriate. The test for significance is a joint test of the parametric form and the parameters of the model. However, the nonparametric approach to subcenter identification developed in this paper ensures that each test is performed on a monocentric subregion of the urban area.

appropriately. A series of tests are undertaken to this end. First, the boundaries are gradually expanded outward from the centroid of the subcenter to include more and more adjacent tracts. Second, a series of concentric rings of increasing sized around the centroid is used to determine the sample included in the monocentric model. The model that maximized R^2 is selected; the t-statistic on the distance variable in this model is the test for subcenter significance.

In all, 30 of the 36 centers were found to be statistically significant, with one additional center being marginally significant. Significance implies that the density within the center was sufficiently more dense that the surrounding area to be considered a employment subcenter. The final results of this exercise – the implementation of the nonparametric procedure for identifying subcenters – is shown in Figure 6. All candidate centers are plotted, but the

Figure 6: Subcenters Identified Using The Nonparametric Approach



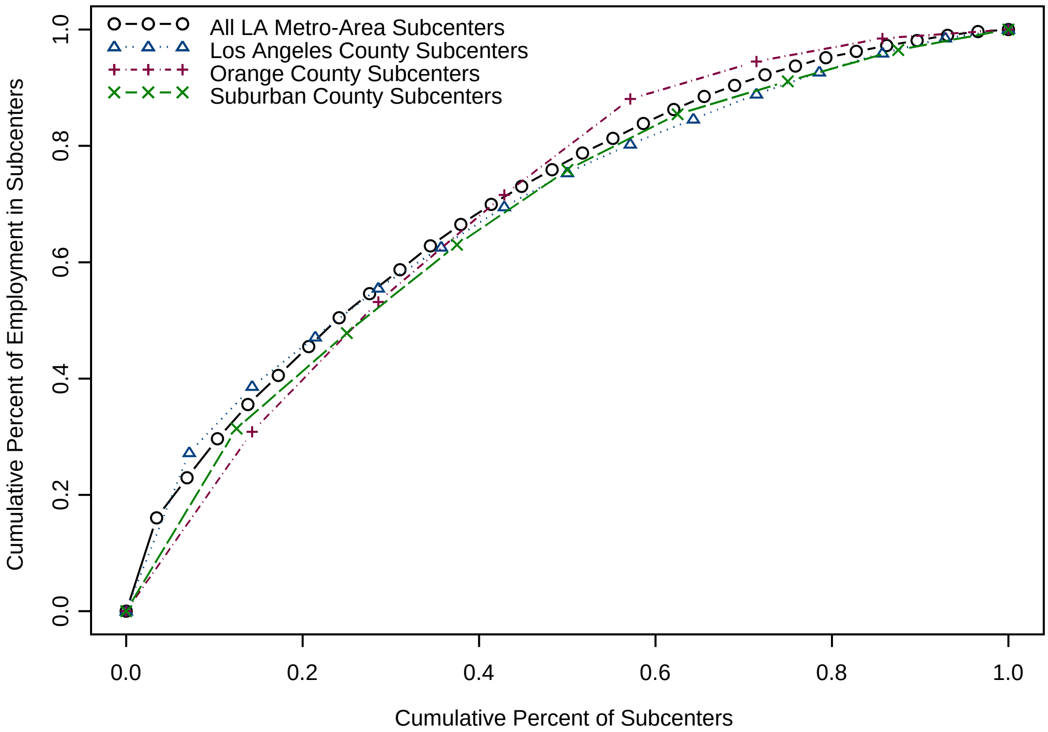
borders are only drawn for subcenters significant at the 5% level.

The nonparametric approach reveals several regularities. First, subcenters are found throughout the CMSA. Agglomeration is not confined to the traditional center or the more familiar large subcenters. Quite the opposite, the lack of employment centers in South Central Los Angeles is the exception not the norm. Second, employment subcenters are smaller in area toward the center of region – although they are comprised of more tracts. This a function of the higher average density in the core of the region. Clusters of office buildings in the core make possible sharply defined subcenters. The sparseness of employment in the far east and west of the region results in more gradual transitions from higher to lower relative density. Another regularity is the clear tracing of the network of highways within the CMSA. This supports urban theory that suggests that subcenter formation should occur at the intersection of major transit nodes to minimize on transit costs.

Interestingly, total employment in the subcenters varies from over 500,000 in the CBD to approximately 12,000 in Laguna Niguel, in the far southern part of Orange County. Figure 7 reveals another regularity regarding the distribution of subcenter employment. For each of four geographies – the entire metropolitan area, Los Angeles County, Orange County, and the combined three outer counties of Riverside, San Bernardino, and Ventura – the distribution of subcenter employment is remarkably consistent. The figure plots the cumulative percentage of employment in subcenters ranked by total employment. Regardless of the geographies, approximately 40% of subcenter employment is found in the top quartile of subcenters; 80% is found in the top half of subcenters.

This relatively even spread of subcenters, if not the employment within them, suggests a steady relationship between the number and relative size of subcenters. That is, there appears to be a hierarchy of subcenters. Again, the notable exception to the relatively uniform spread of subcenters is south-central Los Angeles – on either side of the line tracking south from the city center to Long Beach, where the is no concentration of employment that is distinguishable from it surroundings. The perception of high rates of crime may keep firms

Figure 7: Cumulative Distributions of Subcenter Employment



from locating in South Central Los Angeles and prevent the formation of subcenters in the way that is occurring in the balance of the region.

Several problems are also apparent however. First, the nonparametric approach "finds" a candidate subcenter in Long Beach – where one is expected, but it is not significant. This may be a function of the peculiar physical arrangement of employment there due to the ports, waterways, and rail facilities. The second is the overlap of a handful of subcenters. This follows from the definition of flatness that is part of criteria for boundary definition.

3.2 Alternative Approaches

While the polycentric nature of metropolitan areas has long been recognized, the methods for capturing the complexity have lagged somewhat. This is due in part to the dearth of appropriate and readily-available data, but likely more due to cheap computing power. The last ten years in particular have seen a steady increase in the complexity of regression-based modeling – the nonparametric model proposed in this paper would not have been easy to compute in 1994. The set of methods for subcenter identification developed over the past decade represent steady progress, but each makes some compromise.

3.2.1 Simple Parametric Models

The traditional monocentric, negative-exponential models fail for reasons already discussed, most notably the growing pervasiveness of polycentricity in large urban areas. The natural extension of this model is the polycentric version. In principle, this is an appropriate next step, but it has several drawbacks. The model requires *á priori* specification of the subcenters and the implication of the model is that at every point in the metropolitan area each gradient contributes something to density at that point. This has the effect of making estimation extremely difficult. That is, for use as a comparison to the nonparametric method developed in this paper, numerous specifications were attempted. Convergence was obtained for at most five centers – with the distance to the subcenter being statistically significant for only three candidate centers. Problems with the pre-defined subcenter models makes endogenizing the location of the subcenters impossible.

In light of the results presented for the nonparametric procedure, the polycentric negative-exponential model was not considered more thoroughly. Even in satisfactory subcenter identification were possible, no guidance is typically given for defining the boundaries of the subcenters.

3.2.2 McMillen's Nonparametric Approach

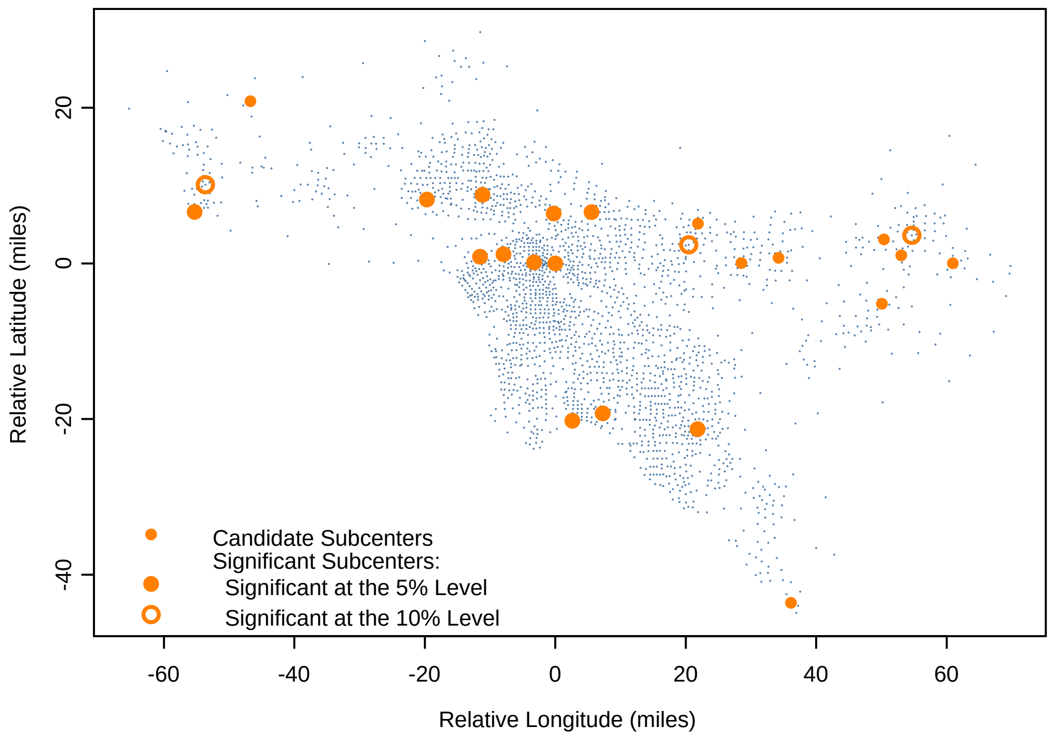
From the fixed-functional form of the parametric models, the trend in modeling the spatial distribution of employment has been towards greater flexibility. The best example of this is McMillen's 2001 model.[11] His approach is also nonparametric, but substantially different in

[11] There are other examples of flexible approaches subcenter definition. A cubic-spline method developed for "satellite city" delineation by Muñiz, Galindo, and García (2003) is analogous to the boundary definition procedure used in this paper – once the density function (population density – in their application) is

its details from the one developed in this paper. His looks for deviations from general spatial trends in density. His first-step locally-weighted regression uses a span of 0.5 – fully half of all observations are used at each point in the grid. Residuals from this smoothed surface are then tested for significance. Those that are significantly greater than the estimated surface are considered candidate subcenters. To test for whether these candidates should be further considered as distinct subcenters, a negative exponential model is estimated for each. As in the case of the nonparametric approach developed in this paper, significance of the subcenter is inferred from the significance on the coefficient on distance variable in the negative exponential model.

Figure 8 shows the estimated the candidate centers derived from McMillen's approach.

Figure 8: Subcenters Identified Using McMillen's (2001) Method



estimated, boundaries are determined by changes in the first or second derivative of the density surface. The authors appear to have some success finding the appropriate satellite cities in Barcelona, Spain. Attempts to replicate this in the context of Los Angeles failed due to the scale of the metropolitan area and the scattered nature of its subcenters. Where Barcelona appears to be monocentric with a ring of satellite cities, the "satellites" around downtown Los Angeles are in highly irregular, and often quite distant, orbits.

It plots 24 tracts where the residuals are significantly greater than the surrounding density surface. The large solid circles indicate the 12 subcenters that are statistically significant (for which the distance from the center is statistically significant in a negative exponential regression of log employment density on distance). An addition three subcenters are marginally significant.

Because of their proximity to one another, they represent at most 12 distinct subcenters. These include the Los Angeles CBD – which appears to be split into two subcenters by the Harbor Freeway (Highway 110), the Wilshire Corridor – also split into two subcenters on either side of the San Diego Freeway (Highway 405), Long Beach, Long Beach Airport, Santa Ana/Irvine, Glendale, Pasadena, the cluster of dense employment in the Sherman Oaks/Van Nuys/Encino area (the intersection of Highways 101 and 405), and Oxnard. Additionally, there is a marginally significant subcenter in the area of Duarte (the intersection of Highways 605 and 210) and in the greater Riverside/San Bernardino area. This set of subcenters is perfectly reasonable, but far from complete. For example, only Santa Ana/Irvine – from among the eight Orange County candidate subcenters identified earlier – is located.

The central problem with this approach is the size of the span parameter used in the local regressions – it pools data points inappropriately. By using 50 percent of the observation at each point, tracts from downtown Los Angeles – some reaching in excess of 300 jobs per acre – are used in regressions meant to determine the height of the density surface in suburban areas in Ventura, Riverside, and San Bernardino, losing local employment surface structure in the process.

Numerous versions of polycentric models were undertaken with little success. McMillen's approach performed the best and was included as the example of "best-practice" among existing regression-based models. Including the poor performance of the mono- and polycentric negative exponential models, none of the regression-based models were successful at capturing the complexity of the distribution of employment in Los Angeles. While McMillen's

(2001) nonparametric approach had some success with U.S. metropolitan areas, neither it nor the others worked well in the context of Los Angeles. Beginning with the monocentric model, each of the others represents progress towards a fully-flexible approach. That said, they do not adequately describe the system of subcenters and are not pursued further.

3.2.3 Clustering Models

Giuliano and Small (1991) develop an alternative to the regression-based models that clusters adjacent tracts to build subcenters based on employment density and total employment thresholds. They define two types of centers: "10-10" and "20-20." The first is an agglomeration of contiguous tracts having a minimum employment density of 10 jobs per acre, and when combined have a total employment of at least 10,000; the second center is analogous with at least 20 jobs per acre and 20,000 total jobs in each subcenter.

Their approach received criticism at first for the apparently arbitrary nature of the density and total employment cutoffs. However, over time, their results held up well against other approaches and what first appeared arbitrary is now called "local knowledge." What the early criticism points to is the fact that the number of subcenters found can be increased by decreasing the threshold and vice versa. "Local knowledge" turned out to be useful as their approach has been replicated numerous times (Anderson and Bogart 2001, Bogart and Ferry 1999, Small and Song 1994).

Giuliano and Redfearn (2004) have updated the original work by Giuliano and Small. In it, the authors replicate the approach using panel data from 1980, 1990, and 2000, in order to track the evolution of the Los Angeles metropolitan area. They find 48 subcenters using the "10-10" definition and 10 subcenters using the "20-20" thresholds for density and total employment, respectively. More importantly, their results closely resemble the employment structure of Los Angeles. Their approach offers no simple statistical test, but the centers they identify are sensible.

4 Subcenters & The Accounting of Employment

Of the regression-based models, the nonparametric approach developed in this paper performed best. That is, the set of subcenters found in the alternative – McMillen's approach – was a subset of less than half of the number of significant subcenters found using the nonparametric approach. And while there is no formal statistical test of the the Giuliano and Small's clustering approach, the results are similar to the nonparametric method in terms of number of subcenters. It is not clear, however, how the two perform relative to each other – from either the spatial distribution of these centers or whether or not there are significant differences in the distributions of employment inside and outside their borders. As a brief examination of the differences between the two methods, this section develops several comparative statistics.

First, Table 2 displays the spatial distribution of the subcenters. It shows that the

Table 2: Spatial Distribution of Centers by County

Model	Los Angeles CMSA	Los Angeles County	Orange County	Other Counties
"20-20"[a]	15	12	3	0
"10-10"[b]	48	38	9	1
Nonparametric	31	15	8	8

a – Centers with an average of 20 jobs per acre and a total of 20,000 jobs.
b – Centers with an average of 10 jobs per acre and a total of 10,000 jobs.

clustering method finds more distinct centers in the denser context of Los Angeles county, where the nonparametric method discovers more of the smaller subcenters found in the three sparser counties. This reveals one idiosyncrasy of the nonparametric approach: that in an area of high average density, fewer centers will be found. The requirement for classification as subcenter is that employment be significantly greater that the surrounding area. This leads to fewer subcenters relative to the "10-10" thresholds in the county with the highest average density – Los Angeles.

Table 3 reports the county shares of employment in and out of centers based the three

Table 3: Share of Employment in Centers by County

Model	Los Angeles CMSA	Los Angeles County	Orange County	Other Counties
"20-20"[a]	16.4	22.1	13.7	0.0
"10-10"[b]	62.5	46.1	43.1	1.1
Nonparametric	51.4	46.7	61.4	56.3

a – Centers with an average of 20 jobs per acre and a total of 20,000 jobs.
b – Centers with an average of 10 jobs per acre and a total of 10,000 jobs.

models. Here, the nonparametric method allocates essentially the same proportion of total employment in Los Angeles County to subcenters as does the "10-10" method. So, while it find fewer centers in Los Angeles, the nonparametric approach reaches the same conclusion regarding concentration of employment in this county. The table also shows that the nonparametric method allocates significantly more of total employment to subcenters in Orange, Riverside, San Bernardino and Ventura counties. In fact, by this method, a greater percentage of employment resides in centers in the suburban counties than in the urban counties.

It is easy to see that the "arbitrary" cutoffs used in the cluster methods is the origin of the differences. By limiting below both density and total employment, the clustering method cannot find the smaller subcenters in the periphery. This is especially true of the "20-20" centers, of which there are none in any of Riverside, San Bernardino, and Ventura, and only three in Orange County.

The "10-10" subcenter cutoffs are more similar to the nonparametric subcenters with regards to both the number of centers and the distribution of employment inside and outside of the centers. The differences are smallest in Los Angeles county, which is the most dense.

However, because the average density is lower in Orange County, and even lower still in the remaining three counties, the "10-10" thresholds become less and less appropriate. They are too high for all but one center in the outlying three counties. The same thresholds that produce reasonable results for Los Angeles County, appear to miss local concentrations of relatively higher employment density that the nonparametric approach does not. This supports the basic motivation for a flexible approach for subcenter identification.

5 Conclusions & Extensions

Polycentricity has become one of the defining characteristics of larger metropolitan areas in the United States. This evolution from single- to multi-centered urban areas necessitates the development of methods that can accurately capture their complexity. This is important for our understanding of urban areas. The growth in interest in spatial mismatch, economies of agglomeration, suburbanization, and many other urban topics begin with a definition of density that may be substantially mismeasured using monocentric- or simple polycentric models.

Monocentric models remain widely used because of their ease of application and because of their historical success. Indeed, there remains cities for which simple parametric forms of employment density may be appropriate. In a large, and growing, number of metropolitan areas, however, a flexible approach capable of capturing highly-irregular spatial distributions of employment is needed. Intermediate approaches such as McMillen's nonparametric approach find some success in other metropolitan areas, but fail to capture the complexity of the system of subcenters in Los Angeles.

The model proposed in this paper is designed specifically to overcome the problems inherent in existing methods of subcenter identification. Using flexible definitions for subcenter location and extent, the subcenters found in the greater Los Angeles metropolitan area are sensible. Moreover, in many cases the subcenter shapes are reflective not of an imposed parametric form, but rather the local infrastructure and natural geography that economic theory predicts as influences of urban form. (It should be noted that, while the application of the nonparametric developed approach here has been to mapping the topography of employment, it could equally well be applied to residential location or other spatially-varying phenomena.)

The nonparametric approach proposed in this paper captures a far-more complex system of subcenters than the other regression-based methods. Identification of a metropolitan area's hierarchy of subcenters offers the opportunity to examine the genesis and evolution of subcenters and to inform our understanding the internal dynamics of urban areas. Moreover, identifying such a system of subcenters provides a rich source of variation for examining a host of urban phenomena. For example, the study of agglomeration depends substantially on the spatial accounting of employment – a task poorly performed by the simpler models. Sprawl is another topic that is not well informed by the parameters in restrictive parametric models. Any urban research that depends on measures of the spatial allocation of employment is likely to be subject to similar problems.

References

Anas, A., R. Arnott, and K. A. Small (1998): "Urban Spatial Structure," *Journal of Economic Literature*, 36(3), 1426–1464.

Anas, A., and I. Kim (1996): "General Equilibrium Models of Polycentric Urban Land Use with Endogenous Congestion and Job Agglomeration," *Journal of Urban Economics*, 40(2), 232–256.

Anderson, N., and W. T. Bogart (2001): "The Structure of Sprawl - Identifying and Characterizing Employment Centers in Polycentric Metropolitan Areas," *American Journal of Economics and Sociology*, 60(1), 147–169.

Bertaud, A., and S. Malpezzi (2003): "The Spatial Distribution of Population in 48 World Cities: Implications for Economies in Transition," Unpublished Manuscript.

Bogart, W. T., and W. Ferry (1999): "Employment Centres in Greater Cleveland: Evidence of Evolution in a Formerly Monocentric City," *Urban Studies*, 36(12), 2099–2110.

Cervero, R., and K.-L. Wu (1997): "Polycentrism, Commuting, and Residential Location in the San Francisco Bay Area," *Environment and Planning A*, 29(5), 865–886.

Cleveland, W. S., and S. J. Devlin (1988): "Locally Weighted Regression: An Approach to Regression Analysis by Local Fitting," *Journal of the American Statistical Association*, 83(403), 596–610.

Garreau, J. (1991): *Edge City: Life on the New Frontier*. New York: Doubleday.

Giuliano, G., and C. L. Redfearn (2004): "Not All Sprawl," Unpublished Mansucript.

Giuliano, G., and K. A. Small (1991): "Subcenters in the Los Angeles Region," *Regional Science and Urban Economics*, 21(2), 163–182.

Gordon, P., and H. W. Richardson (1996): "Beyond Polycentricity: The Dispersed Metropolis, Los Angeles 1970-1990," *Journal of the American Planning Association*, 62(3), 289–295.

Hughes, H. L. (1993): "Metropolitan Structure and the Suburban Heirarchy," *American Sociological Review*, 58(3), 417–433.

McMillen, D. P., and J. F. McDonald (1998): "Suburban Subcenters and Employment Density in Metropolitan Chicago," *Journal of Urban Economics*, 43(2), 157–180.

McMillen, D. P., and S. C. Smith (2003): "The Number of Subcenters in Large Urban Areas," *Journal of Urban Economics*, 53(3), 321–338.

Muñiz, I., A. Galindo, and M. A. García (2003): "Cubic Spline Population Density Functions and Satellite City Deliniation: The Case of Barcelona," *Urban Studies*, 40(7), 1303–1322.

Small, K. A., and S. Song (1994): "Population and Employment Densities - Structure and Change," *Journal of Urban Economics*, 36(3), 292–313.

von Thünen, J. H. (1826): *The Isolated State*. Oxford, New York: Pergammon Press, Translated by Carla M. Wartenberg.