Employment Subcenter Identification: A GIS-Based Method

Qisheng Pan, Li Ma

Abstract

This research studied the methods for identifying employment subcenters and examined the effects of subcenters on surrounding density and housing price. Geographic information system (GIS) was used to organize data and model in a convenient way so that the spatial information such as distance, proximity, and adjacency can be utilized to identify employment subcenters. Houston metropolitan area was selected for the empirical analysis. It used the 1990 Census Transportation Planning Package, the 1990 and 2000 Census summary files 3, and the 2000 employment data obtained from the Houston-Galveston Area Council (H-GAC) to explore subcenters in the Houston area and highlight the changes of the subcenters between 1990 and 2000.

1. Introduction

Most modern metropolitan areas in the U.S. have decentralized employment that has grouped into one or more subcenters outside of Central Business District (CBD) or dispersed in the entire region. The employment subcenters have been described in multicentric models (White 1976; Wieand 1987; Yinger 1992) and nonmonocentric models (Brueckner 1978; Ogawa 1980; Ogawa and Fujita 1980; Fujita and Ogawa 1982). They have also been identified in many U.S. metropolitan areas in empirical studies of Dunphy (1982); Gordon, Richardson, and Wang (1986); Cervero (1989); McDonald and McMillen (1990); Giuliano and Small (1991); McMillen and McDonald (1997, 1998); Cervero and Wu (1998); Craig and Ng (2001); McMillen (2001, 2003); McMillen and Smith (2003).

Two major approaches have been used in subcenter identifications. One is the minimum cutoff point of gross employment density developed by Giuliano and Small (1991) and used by Small and Song (1994), McMillen and McDonald (1998), Cervero and Wu (1997, 1998), and Bogart and Ferry (1999), etc. The other one is the two-stage nonparametric approach proposed by McMillen and McDonald (1997) and revised by McMillen (2001, 2003). Craig and Ng (2001) developed a quantile smoothing splines method, which is also a nonparametric specification in employment density functions but it has not been followed by any other study.

Both the minimum cut-off point method and the two-stage nonparametric approach need to analyze the spatial relationship between objects, such as the adjacency of census tracts. However, the spatial relationship has been identified manually and inefficiently because both models do not have appropriate functions to handle it. Similarly, the empirical studies on the effects of subcenters on surrounding density and housing price also have difficulties in conducting spatial analysis.
Empowered by the spatial functions of GIS, we developed a procedure using Arcview to identify employment subcenters and examine the influences of subcenters on density and housing price of neighboring areas. This procedure is developed on the base of the minimum cut-off point method.

The remainder of this paper was organized as three parts. The next part described a GIS-based method of identifying employment subcenters. After that, it presented an empirical study on employment subcenters in the eight-county Houston region. The last part summarized findings and drew conclusions.

2. Methodology

In this study, we developed a GIS-based procedure to identify employment subcenters and also analyzed the effects of the subcenters on density and housing prices of neighboring tracts.

2.1. The method for identifying employment subcenters

The GIS-based procedure was developed on the base of the minimum cut-off point method proposed and applied by Giuliano and Small (1991). Following the study of McDonald (1987) and McDonald and McMillen (1990), Giuliano and Small (1991) used gross employment density in the formulations of urban subcenters and defined an employment center as a set of neighboring zones, each with density above a cutoff point (D) and total has more than certain number of jobs (J). Their criteria were shown as follows:

\[ D = 10 \text{ jobs/Acre}, \quad J = 10,000 \text{ jobs} \]  

In their study on Los Angeles, Giuliano and Small (1991) found 32 centers, including 28 subcenters and four CBDs in the five-county Los Angeles region in 1980. Similar criteria were used in many other studies. Bogart and Ferry (1999) used the same total minimum employment of 10,000 jobs but a lower employment density of 5,000 jobs per square miles or about 8 jobs per acre to find 9 employment centers in Cleveland, Ohio. Cervero and Wu (1998) changed both the total employment and the density criteria. Total number of jobs is greater than or equal to 95,000 and the density is greater than or equal to 7 jobs per gross acre. They identified 22 employment centers in San Francisco using 1990 CTPP data. Rather than lowering total employment and density criteria as Bogart and Ferry (1999) and Cervero and Wu (1998), McMillen and McDonald (1998) raised the minimum cutoff points to 20,000 total employees and 20 workers per acre to avoid unreasonably large subcenters. They found about 20 employment subcenters in Chicago in 1980 and 1990.

Pointed out by Pan (2003), the minimum cut-off point method has solid statistical meanings. Under the assumption that employment density by tract in a region is distributed normally, the standard score or Z-score values can be used to describe the difference between the density of a certain tract and the average density of all the tracts. \( Z_i \), the Z-score associated with the employment density \( x_i \) in tract \( i \), is given by:
$$Z_i = \frac{x_i - \bar{x}}{\sigma}$$

(2)

where $\bar{x}$ is the mean and $\sigma$ is the standard deviation of the employment density variables.

Given a threshold $T$ on employment density, the above formula is transformed as follows:

$$Z_i = \frac{x_i - \bar{x}}{\sigma} \geq T,$$

(3)

or,

$$x_i \geq \bar{x} + \sigma \times T$$

(4)

A tract is selected when its density is greater than or equal to the mean density plus the multiplication of the density standard deviation and the threshold. Under the assumption that employment density follows normal distribution, to select the tracts that have density higher than 95 percent of tracts, the threshold of $Z$-score value $T$ is set as 1.64. Similarly, $T$ is set as 1.28 to select the tracts that have density higher than 90 percent of the tracts. Therefore, the minimum cutoff density $\bar{x}+1.64 \times T$ can be used to identify the tracts among the 95th percentile and $\bar{x}+1.28 \times T$ is used to identify the tracts among the 90th percentile.

After candidate sites are highlighted as clusters of contiguous tracts with employment density over the minimum density threshold on map, a GIS-based procedure is conducted to identify employment subcenters among the candidate sites using the total employment criteria, i.e. total jobs of a candidate site are greater or equal to a certain number of jobs, e.g. 10,000. The procedure is shown in figure 1.

The first step is to collect employment data for the study region. On Arcview GIS platform, the data includes both a shape file showing the spatial location of tracts and a database file containing job and other attributes of the tracts.

The second step is to calculate job density using an ArcView Avenue program. The [shape] field of a record in feature theme table includes all spatial information of the tract. It is easy to extract area and other attributes from this field to calculate job density.

The third step is to apply density threshold to identify candidate sites. A candidate site is a set of contiguous tracts with job density greater than or equal to a threshold, which is set either by experience, e.g. 10 jobs per acre or using formula (4). The adjacency of tracts is determined using GIS spatial functions.

The last step is to find employment subcenters. It is a simple GIS practice by applying total job criteria to extract subcenters among candidate sites.
Data input
1. Input employment data as Arcview feature map. It contains a shape file showing the geographic locations and the boundaries of tracts and a database file including job information.

Calculate job density
2. Calculate job density of the employment data using an Arcview Avenue program.

Apply density threshold to identify candidate sites
3. Identify candidate sites as sets of contiguous tracts with job density greater than or equal to density threshold. The spatial functions of GIS are used to find the adjacency of tracts.

Apply total job criteria to find employment subcenters
4. Identify employment subcenters among candidate sites using the total job threshold, i.e. 10,000 jobs.

Figure 1. a GIS-based procedure to identify employment subcenters.
2.2 The method for analyzing the effects of employment subcenters

Eric Heikkila et al. (1989) developed a functional form to transfer the formulations of the accessibility component of land values from monocentric urban form to polycentric structure, which is generalized as follows.

\[ A_m = B \prod_n f(d_{mn}) \]  \hspace{1cm} (5)

where \( A_m \) is the accessibility of land values at zone \( m \)
\( B \) is a matrix of parameters
\( d_{mn} \) is the distance between zone \( m \) and zone \( n \)
\( f(d_{mn}) \) is the impedance function.

The formulation of \( f(d_{mn}) \) was given as \( \frac{1}{d_{mn}} \). Anas and Small (1998) suggested that impedance function as a negative exponential function as follows:

\[ f(d_{mn}) = \exp(-\alpha_n d_{mn}) \]  \hspace{1cm} (6)

where \( \alpha_n \) is coefficient of the negative exponential function
\( d_{mn} \) and \( f(d_{mn}) \) are the same as previously described.

To estimate the effects of subcenters, we transformed the equation (5) to the following formulation by applying natural logarithm to both sides:

\[ \ln A_m = \sum_n b_n \ln(f(d_{mn})) \]  \hspace{1cm} (7)

where \( A_m \), \( b_n \), \( d_{mn} \), and \( f(d_{mn}) \) are the same as previously described.

A negative exponential formulation \( f(d_{mn}) = \exp(-\alpha_n d_{mn}) \) was utilized for function \( f(d_{mn}) \) and the equation (7) was converted as follows:

\[ \ln A_m = \sum_n b_n' d_{mn} \]  \hspace{1cm} (8)

where \( b_n' \) is a parameter as the product of \( b_n \) and -\( \alpha_n \).
\( A_m \) and \( d_{mn} \) are the same as previously described.

It showed that the natural algorithm of density or housing price is modeled as dependent variables and the distances to CBD and each employment subcenter are incorporated as independent variables in the linear function. The distance variable is calculated using spatial functions in GIS. The regression analysis was conducted on statistical package, such as SPSS.
3. Empirical Study

The GIS-based procedure was implemented to study employment subcenters in Houston metropolitan area. Employment data came from the 1990 CTPP and the 2025 regional growth forecast study by H-GAC. The 1990 CTPP provided employment data for 2194 TAZs in the eight-county Houston region, which is Montgomery, Liberty, Waller, Harris, Chambers, Fort Bend, Brazoria, and Galveston. The 2000 H-GAC data included employment for 2634 TAZs in the H-GAC region, which is the same as the eight-county Houston region. The median housing price information was extracted from Census summary file 3 (SF3) at block group level and plugged into the employment data by TAZ in 1990 and 2000 using a spatial analysis tool.

3.1. The employment subcenters in the H-GAC region

We applied the GIS-based procedure described in Figure 1 to identify employment subcenters in Houston in 1990 and 2000.

At the beginning, we tested the distribution of job density and we got confirmation that the natural logarithm values of job density, i.e. Ln(job density), in both 1990 and 2000 follow normal distribution.

Then, we explored the relationship between the minimum cutoff points and the Z-score values of Ln(employment density). We found that minimum employment density of 10 employees per acre, which was used as minimum density point by Giuliano and Small (1991), was corresponding to the Ln(employment density) cutoff point of 2.30 employee per acre or Z-score of about 1.16.

As we know, Z-Score value of 1.16 is corresponding to the 87.7% probability in one-tail hypothesis test. It is clear that the density cutoff point proposed by Giuliano and Small (1991) is to find the TAZs with Ln(density) values higher than 87.7% of regional TAZs.

When we utilized the criteria with the minimum Z-Score value of 1.16 to employment by TAZ, we selected 162 TAZs from 1990 CTPP and 175 TAZs from 2000 H-GAC data. After applying the minimum total employment of 10,000 to the data, we identified 11 employment centers compromising 130 TAZs (Figure 2) in 1990 and 12 employment centers consisting of 127 TAZs in 2000 (Figure 3). Table 1 showed the difference between the 11 employment centers in 1990 and 12 employment centers in 2000.
Figure 2. Houston Employment Subcenters, 1990, identified by Z-score greater than 1.16.

Figure 3. Houston Employment Subcenters, 2000, identified by Z-score greater than 1.16.
Table 1. Comparison of Houston employment centers in 1990 and 2000, identified by the Z-score cutoff point method

<table>
<thead>
<tr>
<th>ID</th>
<th>Employment centers in 1990</th>
<th>Employment centers in 2000</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>name</td>
<td>Total employees</td>
</tr>
<tr>
<td>1</td>
<td>DOWNTOWN</td>
<td>155,268</td>
</tr>
<tr>
<td>2</td>
<td>NASA</td>
<td>12,241</td>
</tr>
<tr>
<td>3</td>
<td>GREENSPOINT</td>
<td>13,412</td>
</tr>
<tr>
<td>4</td>
<td>NORTHWEST MALL</td>
<td>26,936</td>
</tr>
<tr>
<td>5</td>
<td>BARKERS LANDING</td>
<td>16,068</td>
</tr>
<tr>
<td>6</td>
<td>WEST CHASE</td>
<td>15,569</td>
</tr>
<tr>
<td>7</td>
<td>SHARPSTOWN</td>
<td>18,488</td>
</tr>
<tr>
<td>8</td>
<td>GALLERIA</td>
<td>79,875</td>
</tr>
<tr>
<td>9</td>
<td>MONTROSE</td>
<td>15,587</td>
</tr>
<tr>
<td>10</td>
<td>WESTHEIMER AND MEDICAL CENTER</td>
<td>116,616</td>
</tr>
<tr>
<td>11</td>
<td>-----</td>
<td>-----</td>
</tr>
<tr>
<td>12</td>
<td>GALVESTON</td>
<td>15,614</td>
</tr>
<tr>
<td>13</td>
<td>-----</td>
<td>(8,048)</td>
</tr>
<tr>
<td>Others</td>
<td>-----</td>
<td>1,302,183</td>
</tr>
<tr>
<td>Total</td>
<td>-----</td>
<td>1,787,857</td>
</tr>
</tbody>
</table>

Source: Author calculation using 1990 CTPP data and 2000 H-GAC employment data
3.2. The effects of employment subcenters on surrounding density and housing prices

The accessibility to employment centers was hypothesized to have influence on surrounding employment, population and housing price. Based on formula (8), we ran multivariate regression to analyze the change of employment density, population density, and housing price with respect to distance to employment subcenters in both 1990 and 2000. Table 2 presented the results of regression analysis.

Table 2 showed that the a large percent of the variance of employment density, population density, and median housing price can be explained by the distance to the subcenters, for example, 48 percent of the variance of Ln(employment density), 33.2 percent of the variance of Ln(Population Density), and 37.6 percent of the variance of Ln(Median Housing Price) can be explained by the distance to the 11 subcenters in 1990. Similarly, 36 percent of the variance of Ln(employment density), 23.7 percent of the variance of Ln(Population Density), and 41.8 percent of the variance of Ln(Median Housing Price)  can be explained by the distance to the 12 subcenters in 2000.

Table 2 also indicated that the relationship between the dependent variable and the independent variable is significant because of the overall p-value of 0.000 crossing all the columns.

It confirmed the traditional relationship between density, housing price and distance in a monocentric model because Ln(employment density) and Ln(population density) decrease with the increasing distance to Downtown and some subcenters, which was represented by negative regression coefficients in Table 2. The p-value of almost 0.000 also showed that these centers have played a dominant role on the surrounding density and housing price. The positive coefficients in some relatively small subcenters, such as NASA, Greenspoint, Northwest Mall, etc. indicated that the logarithm values of employment density and population density increase with the distance to the centers, which means that employment density around the subcenters has been significantly influenced by CBD or other subcenters.

An interesting comparison between 1990 and 2000 shows that the distance to subcenters has had less explanatory power on density but it is more influential on housing price. For example, the adjusted $R^2$ of Ln(Employment Density) and Ln(Population Density) decreased from 0.480 to 0.36 and from 0.332 to 0.237 respectively while the adjusted $R^2$ of Ln(Median Housing Price) increased from 0.376 to 0.418. It implied that CBD and existing subcenters have less effect on job density and population density. In other words, employment and population has become more decentralized in the last decade. However, CBD and subcenters have become more influential in housing price.
Table 2. Regression analysis of the impacts of identified subcenters, 1990 and 2000

<table>
<thead>
<tr>
<th>Variable (Distance to the centers)</th>
<th>1990 CTPP data (Sample size = 2194)</th>
<th>2000 H-GAC data (Sample size = 2455)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Ln(Employment Density)</td>
<td>Ln(Population Density)</td>
</tr>
<tr>
<td>Constant</td>
<td>3.391550144 (t = 8.386, p = 0.000)</td>
<td>3.837241892 (t = 12.444, p = 0.000)</td>
</tr>
<tr>
<td>DOWNTOWN</td>
<td>-7.0409E-05 (t = -2.526, p = 0.012)</td>
<td>-2.82108E-05 (t = -1.310, p = 0.190)</td>
</tr>
<tr>
<td>GALVESTON</td>
<td>-2.3633E-05 (t = -15.435, p = 0.000)</td>
<td>-1.40474E-05 (t = -11.964, p = 0.000)</td>
</tr>
<tr>
<td>NASA</td>
<td>3.05732E-05 (t = -12.699, p = 0.000)</td>
<td>1.46483E-05 (t = 7.754, p = 0.000)</td>
</tr>
<tr>
<td>GREENSPOINT</td>
<td>7.63335E-06 (t = 2.183, p = 0.029)</td>
<td>-6.53698E-07 (t = -0.243, p = 0.808)</td>
</tr>
<tr>
<td>NORTHWEST MALL</td>
<td>8.90654E-05 (t = 3.913, p = 0.000)</td>
<td>-5.42203E-06 (t = -0.311, p = 0.756)</td>
</tr>
<tr>
<td>BARKERS LANDING</td>
<td>1.76117E-05 (t = 1.809, p = 0.071)</td>
<td>2.09559E-05 (t = 2.833, p = 0.005)</td>
</tr>
<tr>
<td>WEST CHASE</td>
<td>-5.2464E-05 (t = -1.533, p = 0.125)</td>
<td>-4.90050E-05 (t = -1.902, p = 0.057)</td>
</tr>
<tr>
<td>SHARPSTOWN</td>
<td>9.91495E-05 (t = 2.152, p = 0.031)</td>
<td>3.66819E-05 (t = 1.059, p = 0.290)</td>
</tr>
<tr>
<td>GALLERIA</td>
<td>-0.00031976 (t = -6.967, p = 0.000)</td>
<td>-3.76931E-05 (t = -1.087, p = 0.277)</td>
</tr>
<tr>
<td>WESTHEIMER</td>
<td>0.000169992 (t = 3.582, p = 0.000)</td>
<td>1.73433E-05 (t = 0.479, p = 0.632)</td>
</tr>
<tr>
<td>MONTROSE</td>
<td>2.66369E-05 (t = 0.676, p = 0.499)</td>
<td>2.69861E-05 (t = 0.891, p = 0.373)</td>
</tr>
<tr>
<td>MEDICAL CENTER</td>
<td>----</td>
<td>----</td>
</tr>
<tr>
<td>FAIRBANKS</td>
<td>----</td>
<td>----</td>
</tr>
<tr>
<td>R² (adjusted)</td>
<td>0.480</td>
<td>0.332</td>
</tr>
<tr>
<td>Standard error of the estimate</td>
<td>2.003</td>
<td>1.489</td>
</tr>
<tr>
<td>F-test</td>
<td>185.169</td>
<td>93.530</td>
</tr>
<tr>
<td>p-value</td>
<td>0.000</td>
<td>0.000</td>
</tr>
</tbody>
</table>

Source: Author calculation
4. Conclusions and discussions

This paper discussed the methods for identifying employment subcenters and analyzing the effects of CBD and the subcenters. A GIS-based procedure was developed to incorporate the spatial analysis functions of GIS to a Z-score cutoff point method, which provided standard criteria to various sizes of sample data and different study areas.

We implemented the method to find employment subcenters in the eight-county Houston region in 1990 and 2000. Data sources included the 1990 CTPP Part 2, the 2000 H-GAC employment data set, and the Census summary file 3 in 1990 and 2000. We identified 11 employment centers in 1990 and 12 centers in 2000.

We also examined the effects of the identified subcenters on regional employment density, population density and housing price. We found that a large portion of the variance of density and housing price can be explained by the accessibility to CBD and subcenters and also the relationship between the density or housing price and the accessibility to the centers are significant in most cases. CBD and subcenters played different role in density and housing price gradient, for example, employment density and population density decrease with the distance to Downtown Houston because of the negative coefficient resulted from regression analysis while housing price increase with the distance to Downtown Houston because of the positive coefficient.

The comparison of the results between 1990 and 2000 showed that downtown still plays an important role in the density and housing price gradient but its power has been downgraded due to the continuing decentralization and the emergence of more employment centers. The complementary effects of CBD on density and housing price have been confirmed.

Reference


Yinger J. (1992) “City and suburb: urban models with more than one employment center”
Contact Information
Primary Author
Dr. Qisheng Pan
Texas Southern University
School of Public Affairs
HH 334E, 3100 Cleburne Ave.
Houston, TX 77004
US
Tel: 713-313-7221
Fax: 713-313-7447
Email: pan_qs@tsu.edu

Co-Author
Ms. Li Ma
Texas Southern University
Department of Computer Science and Physics
NSB 126, 3100 Cleburne Ave.
Houston, TX 77004
US
Tel: 713-313-7028
Fax: 713-313-7447
Email: malx@tsu.edu