Using a GIS for Real Estate Market Analysis: The Problem of Spatially Aggregated Data

*John Clapp**
*Mauricio Rodriguez***

Abstract. **Many databases used for real estate market analysis are not available at the address level. For example, information on employment and unemployment may be available only for labor market areas; and Census data is typically tabulated for blocks or higher levels of spatial aggregation. A Geographic Information System (GIS) associates these spatially aggregated data with the geographical center of the area. This poses special problems when we use a GIS to evaluate linkages between supply and demand. This article presents some solutions to this problem; methods that are relatively easy to implement on a GIS are emphasized. A GIS can be used to calculate a theoretical average travel distance to the population in the geographical area. We propose ways to determine when these theoretical distances are inadequate approximations; and we provide alternatives for these situations.**

Introduction

A fundamental purpose of real estate market analysis is to quantify the relationship between supply and demand for a given property type in a local market. A Geographical Information System (GIS) is a powerful tool for storing and manipulating large amounts of information on spatial relationships between supply and demand.[1] But, most GIS software is generic, not designed for measurement of linkages identified by urban economic theory. This article is motivated by the need to guide future real estate analysis that makes use of widely available data within a GIS framework. Urban economic and real estate applications require the measurement of certain types of linkages. How good is GIS at providing this information? How can linkage information best be analyzed?

GIS software is well designed for manipulating address level data. However, a lot of important information is not available at the address level. For example, Census data are available for city blocks, Census tracts or higher levels of aggregation, with more detailed data available for larger spatial units. Higher levels of spatial aggregation are employed for more thinly populated areas. Other important information, such as labor market data, are typically available at high levels of spatial aggregation.[2] Thus, spatial data are often grainy.

*University of Connecticut, Storrs, CT 06269 or jmclapp@uconnvm.uconn.edu.
**Texas Christian University, Fort Worth, TX 76129 or mrodriguez@zeta.is.tcu.edu.

Distance measurements between points of interest have been used in a variety of real estate analyses.[3] Although some of the cited research makes use of the technological advantages provided by GIS, address specific data is required to implement the analysis. Since these data are scarce, spatially aggregated data are often used in real estate research.[4] There are many situations where it may be of interest to examine distance measurements that involve grainy data. Anytime the location of the property under study (such as a retail, office or apartment building) is in the same geographic area as the area for which spatially aggregated data are available (same zip code, Census tract, town, etc.), distance measurements may approach zero. Distance measurements will approach zero as the location of the property under study gets closer to the center of the geographic area. This will occur regardless of the true spatial distribution of the address data or the true relationship between supply and demand for the property being analyzed.

A GIS handles spatially aggregated data by associating it with the centroid of the Census block group, Census tract or other geographic area.[5] This method poses problems when quantifying linkages between supply and demand. For example, a store near a centroid will have near zero estimated distance to demand in that area. But, actual distance traveled will be significantly greater than zero, since the population does not live in a high-rise building at the centroid. This poses significant problems for market analysis, especially since distance, which makes a negative contribution to potential demand, often enters in the denominator of typical calculations.

We use the gravity model to investigate a general class of problems associated with spatial aggregation.[6] We show how a GIS can correct for these problems by substituting a travel distance that corresponds more closely to reality.

The next section presents theoretical results on average distance to the centroid of a region that takes on various shapes (circle, square, etc.). We show how theoretical travel distances can be substituted for travel distances calculated under the assumption that all population or employment is located at the centroid. The third section uses a gravity model with a GIS to show how the conclusions from market analysis are changed by using theoretical distances rather than distances to centroids. Both the second and third sections emphasize practical approaches to diagnosing and curing problems with spatially aggregated data. The fourth section is the conclusion.

Theoretical Average Travel Distances

A GIS excels in measuring linkages by either straight line or road distances.[7] But, spatial aggregation puts population, income, employees and the like at the centroids of geographical areas. In essence, one must assume that all of the population or employment is housed in a high-rise building at the geographical centroid.[8] The problem we address is how to calculate distances from one activity (*e.g.*, a store or shopping center) to the data (*e.g.*, population or employment) that the GIS associates with geographical centroids. Thus, we seek to work with this simplification imposed by spatially aggregated data.

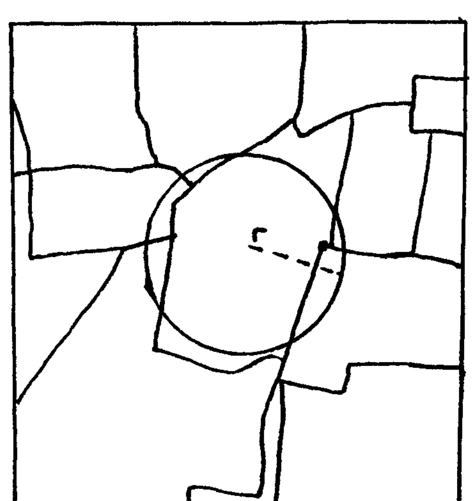
This simplifying assumption works well when the store is a long way from the center of the tract (or other geographical area). But, what happens when the store is close to the population center of the tract? In this case, the traditional analysis (with or without making use of a GIS) typically assumes that there is virtually no distance for customers to travel to the store. But this is untrue since some customers still have to travel from the far corners of the Census tract. Some adjustment must be made when the store is close to the center of the geographic area under study.

The Theoretical Radius

The first adjustment we proposed is based on the concept of the theoretical radius. First, we draw a circle centered on the center of the tract and with an area equal to the area of the tract (see Exhibit 1). The radius of this circle is an approximation to the distance that a customer at the boundary of the tract would have to travel in order to reach a store somewhere near the center of the tract. The radius under the assumption that the Census tract is a circle is an approximation to the average distance that must be traveled.[9]

The theoretical radius is easily and quickly calculated with information available in any GIS software; namely, information on the area of each polygon (see Exhibit 1 for the formula). We propose that the GIS can be used to substitute two-thirds times the theoretical radius whenever actual distance to the centroid is less than two-thirds

Exhibit 1
The Theoretical Radius



x = Geographical Centroid
A = Area of the Census tract
r = Theoretical Radius = $[A/\pi]^{.5}$

of the theoretical radius. Two-thirds times the theoretical radius is the average travel distance for a population uniformly distributed on a circle (see the Appendix for a derivation).[10]

Actual observation of Census tract boundaries indicates that the most common geometric shape is square, followed by a rectangle and irregular shapes. Does the theoretical radius work if the shape of the Census tract is very different from that of a circle? That is, at what point does the theoretical radius cease to be a good approximation? And, in these cases, what method should be used to calculate distances?

Accuracy of the Theoretical Radius

The theoretical radius is an approximation, simple to calculate with a GIS. When does the use of the theoretical radius (or, more precisely, two-thirds of the theoretical radius) introduce inaccuracies into the analysis?

To answer this question, it is useful to analyze geographical regions with shapes that differ from a circle. We proceed to examine regions that are shaped like a square or a rectangle, as well as irregular shapes. Average travel distance for each of these shapes is compared to that of a circle with the same area (*i.e.*, the theoretical radius concept).

Equations (1), (2) and (3) summarize results for average travel distance, T_i, where $i = c$ (circle), s (square) and R (rectangle). Details on the derivation of these formulas are contained in the Appendix.

$$T_c = \frac{2}{3} r, \tag{1}$$

where:

$$\begin{aligned} r &= \text{Radius of the circle} \\ &= \left(\frac{A}{\pi}\right)^{.5}. \end{aligned}$$

$$T_s = \frac{1}{3} xs = .3826x, \tag{2}$$

where:

x = Length of side of the square
s = Integration constant
= 1.1478.

$$
\begin{aligned}
T_R &= k(z)\, X_1 \\
&= .5932\, X_1 \quad \text{if } z = 2,
\end{aligned}
\tag{3}
$$

where:

X_1 = Length of the short side of the rectangle,
X_2 = Length of the other (long) side,
z = X_2/X_1 and
$k(z)$ = A constant for each z, tabulated in the Appendix.

No closed-form solution was found for a triangle; the integral in this case involves the square root of a sum of squares. However, a triangle, like irregular shapes, can be examined through numerical integration as discussed below.

Now we can examine the main question of this section by holding area constant (*i.e.*, normalizing on the size of the geographical area) and examining the ratio of actual travel distance to theoretical distance assuming that the geographical region is shaped like a circle. We begin with a region that is shaped approximately like a square:

$$
r^* = \frac{x}{\sqrt{\pi}} = .5642x, \tag{4}
$$

where:

r^* = The theoretical radius for a circle with area equal to the square with side x.

$$
\frac{T_s}{T_c} = \frac{.3826x}{\frac{2}{3}(.5642)x} = 1.0172. \tag{5}
$$

Equation (5) shows that if the local area is a square, the average travel distance will be 1.72% greater than a circle with the same area: the theoretical radius will give a result that is too low by 1.72%.

If the actual shape is approximately a rectangle:

$$
r^* = \frac{1}{\sqrt{\pi}}(X_1 X_2)^{.5} = .5642(X_1 X_2)^{.5}. \tag{6}
$$

where:

r^* = The theoretical radius for a circle with area equal to the rectangle with sides X_1 and X_2.

$$\frac{T_R}{T_c} = k(z)\, X_1/(.6667 \times .5642\, X_1 X_2)^{.5}$$

$$= \frac{.5932}{.5319} = 1.1152 \quad \text{if } z = 2. \tag{7}$$

Equation (7) shows that if $X_1 = .5X_2$ (one side of the rectangle is half the length of the other side, $z = 2$), then $T_R/T_c = 1.1152$, so the average travel distance from using the theoretical radius will be low by a factor of only 11.52%. If $z = 3$, the factor increases to 26.35%; $z = 4$ implies 41.02%.

To summarize, the theoretical radius provides a good approximation if the shape of the geographical area can be approximated by a circle or by a square. But, as a rectangular shape is distorted away from the shape of the square, the approximation becomes increasingly poor. Intuitively, this follows because people in the far ends of a rectangle have a long travel distance to get to the centroid; these long distances begin to dominate average travel distance as the rectangle is elongated. The same is not true of a circle since the circle keeps everyone as close as possible to the center.

Irregular Geographical Regions

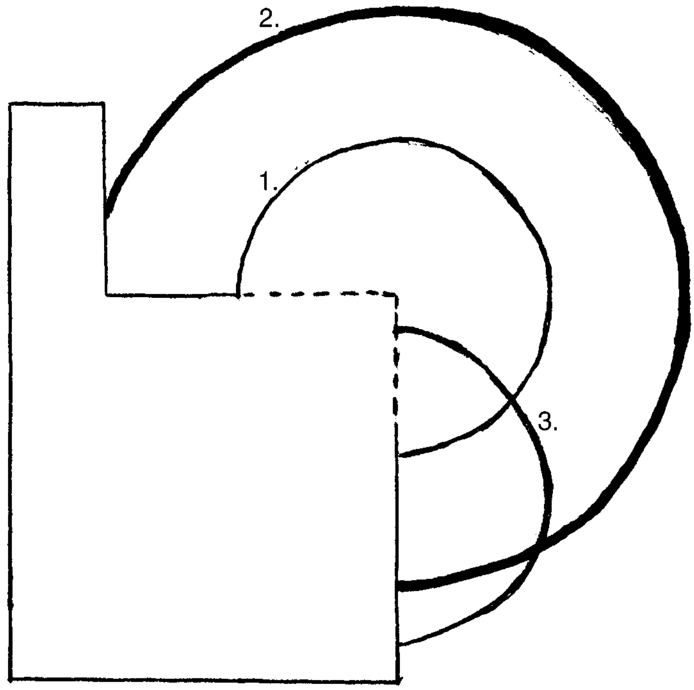
We evaluate average travel distance on irregular geographical regions in order to describe the robustness of the theoretical radius. This section uses numerical integration to find average travel distance on irregular regions given by Exhibit 2.[11] The irregular shapes shown in Exhibit 2 were designed to be typical of those shapes we observed in Census tracts. These shapes are often given by rivers or other geographical barriers.

Exhibit 3 compares the average travel distance for the irregular shapes to the average travel distance based on the theoretical radius. We conclude that a circle is a good approximation (for purposes of average travel distance) to irregular but bulky shapes. Even when these shapes have a significant rectangular protrusion, the circle does well as long as a significant part of the shape is not rectangular. Thus, the problem with the elongated rectangular shape (*e.g.*, $z = 3$ or 4) is a rather special one that occurs only when the rectangular shape dominates.[12]

Practical Identification of a Poor Approximation

We have determined that the theoretical radius (multiplied by two-thirds) is a poor approximation to average travel distance on a rectangle with one side three or four times (or more) the length of the other ($z = 3$ or 4). We would like to find some way

Exhibit 2
Irregular Shapes Evaluated Numerically



to use a GIS to quickly identify geographical shapes that cause the theoretical radius to be a poor approximation.

Before proceeding, however, we should note that a poor approximation to average travel distance is different than a poor approximation for the purposes of any particular

Exhibit 3
Average Travel Distances in Irregularly Shaped Regions

	Irregular Shape			Comparison		
	Area	Perim.	Avg. Travel Distance	T_c	Ratio	R Perim.
Shape 1	656	121.7	10.52	9.63	1.092	1.341
Shape 3	597	109.1	9.93	9.19	1.081	1.260
Shape 2	1,023	140.7	12.38	12.00	1.032	1.244

Notes: T_c is the average travel distance for a circle with the same area as the irregular shape. Ratio is the ratio of the two average travel distances; shapes are ordered by decreasing values for Ratio. R Perim. is the ratio of perimeters: Irregular shape divided by the circle.

market analysis. This is because the theoretical radius avoids the gross error associated with using small (possibly near zero) distances to the centroid. In particular, market analysis may be insensitive to the remaining error, even if the average travel distance is understated by an order of magnitude. Our numerical example in the next section illustrates such a situation.

We investigated two methods for identifying geographical regions whose shape cannot be adequately approximated by a circle. The first of these, average node distance, uses all of the points (nodes) along the boundaries of the geographical area. These nodes enable the GIS to draw the boundaries of the region by connecting the nodes with straight lines. The method we investigated compares the average distance from these nodes to the centroid to the theoretical radius. The details of this exercise are not reported because we found a more convenient way to identify the shapes in question.

A convenient method simply takes the ratio of the perimeter of the geographical region to the perimeter of a circle with the same area (*i.e.*, the perimeter associated with the theoretical radius). This ratio is very easy to calculate since most GIS software provides the perimeters of all polygons; we have already determined that the theoretical radius is convenient to calculate from the area of the polygon.

To evaluate the performance of this ratio of perimeters, we first look at analytical results for a square and a rectangle, both as a ratio to a circle with the same area:[13]

$$\frac{P_s}{P_c} = \frac{4x}{2\pi r^*} = \frac{4x}{2\pi(.5642)x} = 1.1284. \tag{8}$$

Similar calculations for a rectangle show that:

$$\begin{aligned}\frac{P_R}{P_c} = \frac{2(X_1 + X_2)}{2\pi r^*} &= \frac{(X_1 + X_2)}{\pi(.5642)(X_1X_2)^{.5}} \\ &= .5642(X_1 + X_2)(X_1X_2)^{-.5}.\end{aligned} \tag{9}$$

If $X_1 = .5X_2$ $(z = 2)$, then $P_R/P_c = 1.1968$. If $z = 3$, $P_R/P_c = 1.3029$; $z = 4$ implies $P_R/P_c = 1.4105$.

When Equations (8) and (9) are compared to Equations (5) and (7), it is apparent that the ratio of perimeters moves in the right direction. That is, when the ratio of perimeters increases, the ratio of average travel distances also increases. But the ratio of perimeters is not as sensitive as the ratio of travel distances because the perimeter does not give us much weight to points that are distant from the centroid.

For irregular shapes, the perimeter ratio is summarized in Exhibit 3. This confirms the conclusion from Equations (8) and (9).

The only problem with the ratio of perimeters is that it is not as sensitive as the ratio of average travel distances. This leaves us with an empirical question: Is the ratio of perimeters sensitive enough to identify geographical shapes that are not readily approximated by a circle? We examine this question in the next section.

Alternative Measures of Linkages. If the theoretical radius is determined to be a poor approximation for a given analytical purpose, what alternative measures of linkages are available? For example, when the ratio of perimeters, or other information, says that the actual travel distance is much greater than two-thirds of the theoretical radius, we need an alternative measure of linkages that is not too difficult to calculate with a GIS. Of course, we would use these alternatives only when the outcome of a particular model is changed significantly by errors in the measurement of linkages.

Alternative measures of linkages, arranged roughly in order of increasing ease of application are:

1. Graph the shape and use a numerical approximation to average travel distance.
2. Estimate the average travel distance along roads by using a GIS to calculate travel distance to the centroid from a random sample of nodes along roads within the subject geographical area.
3. Examine the shape of the geographical region in question to see if it can be approximated by a square or a rectangle. If so, the formulas provided here can be used.

The ease with which the second method can be implemented depends on the functionality built into a GIS. It may require some programming effort within a GIS.

A Numerical Example Using Gravity Analysis

In this section we use a GIS and gravity analysis to examine if the conclusions from market analysis are changed by using theoretical distances rather than traditionally used distances to centroids. Specifically, we examine changes in market shares estimated for retail sites.

The gravity spatial interaction model has been a popular urban model for many years. The gravity model assumes that the attraction between consumers in a geographic area and retail sites is proportional to their sizes and the impedance to interaction between these entities is directly related to the cost of traveling the distance between them. We use a single-constrained gravity spatial model of the form developed by Huff (1963) and explained by Haynes and Fotheringham (1984). The model can be used to estimate market shares for retail sites from a trade area or from specific neighborhoods. Equation (10) estimates the attractiveness for spatial interaction, A_{ij}, between neighborhood i (or other level of geographic area) and retail site j using an inverse power function in which the attractiveness of each retail site, S_j, is raised to an appropriate power β, and the impedance to interaction (distance) grows by the power of λ.

$$A_{ij} = S_j^{\beta} D_{ij}^{-\lambda}. \tag{10}$$

The market share captured by retail center j from the consumers of neighborhood i, p_{ij}, is estimated by dividing the attractiveness of retail site j to neighborhood i, by the sum of attractiveness measures for all competing retail sites in the trade area under study. Equation (11) shows how to estimate market shares described earlier.[14]

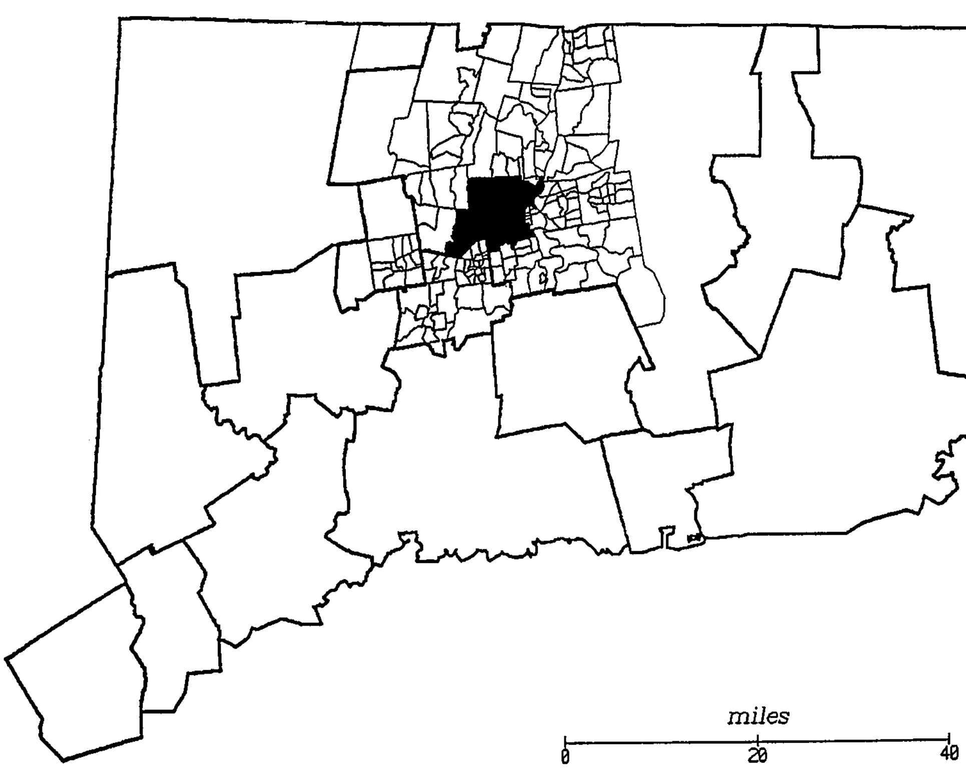
$$p_{ij} = S_j^{\beta} D_{ij}^{-\lambda} / \sum_k S_k^{\beta} D_{ij}^{-\lambda}. \tag{11}$$

The market share estimates represent the probability that consumers from neighborhood i will shop in retail center j.

The Data

We study a trade area of grocery stores in West Hartford, Connecticut. The location of the trade area within the state of Connecticut is shown in Exhibit 4. The average number of households within the sixty-three Census tracts studied is 1,180. The largest number of households within a tract is 2,958, and the smallest number of households

Exhibit 4
Grocery Store Trade Area in West Hartford, CT



within a tract is 2. The average median household income across all of these tracts is $32,789. The largest median household income within a tract is $95,340, and the smallest median household income within a tract is $6,242.

Estimated Market Shares

Our initial analysis estimates market shares for six existing supermarkets.[15] First, we estimate a gravity model ignoring the potential pitfalls of traditionally used distance measures. The trade area and location of the stores under study are shown in Exhibit 5. Next, we estimate market shares for the same six grocery stores, but we make use of the theoretical radius provided by a GIS. At first glance there appears to be little difference between the two estimates. The overall market share captured by each store does not significantly change. However, the estimated market share for the Census tract(s) that make use of the theoretical radius may be notably different than those that do not. Exhibit 6 shows the difference in estimated market shares for a Census tract where the theoretical radius is employed. The initial market share estimates for stores 5 and 6 are 40.8% and 39.1%, respectively, for Census tract 090034974. The theoretical radius provides market share estimates for stores 5 and 6 equal to 33.4% and 43.2%, respectively, in the same Census tract. Overestimation (over 7% market share for store 5) and underestimation (over 4% market share for store 6) could translate into significant misallocated dollar amounts for advertising and expansion.

Exhibit 5
Location of Grocery Stores in Trade Area

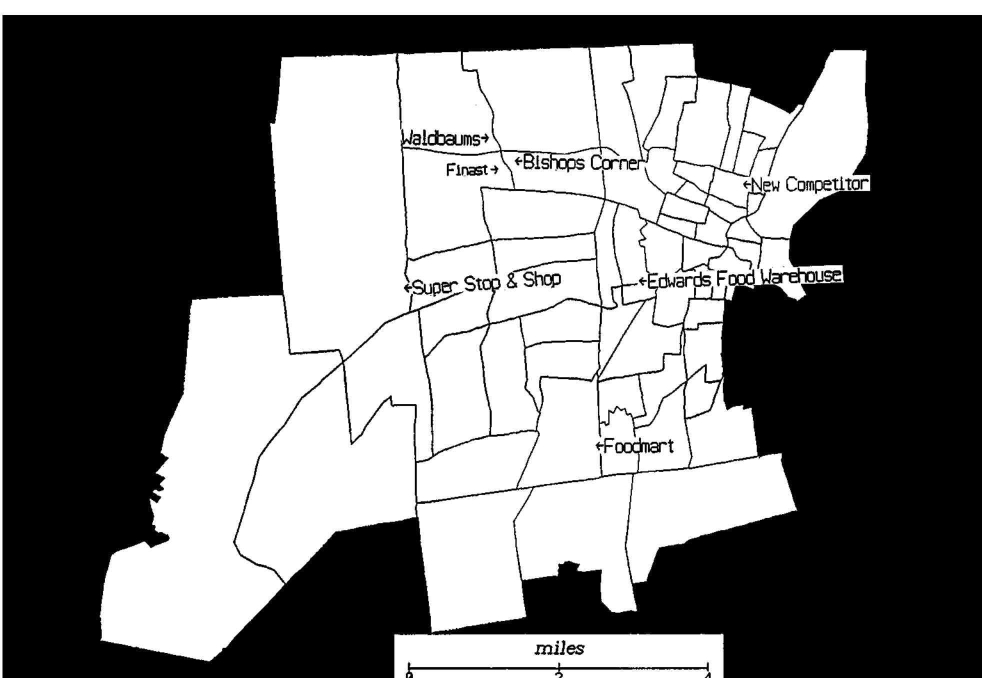
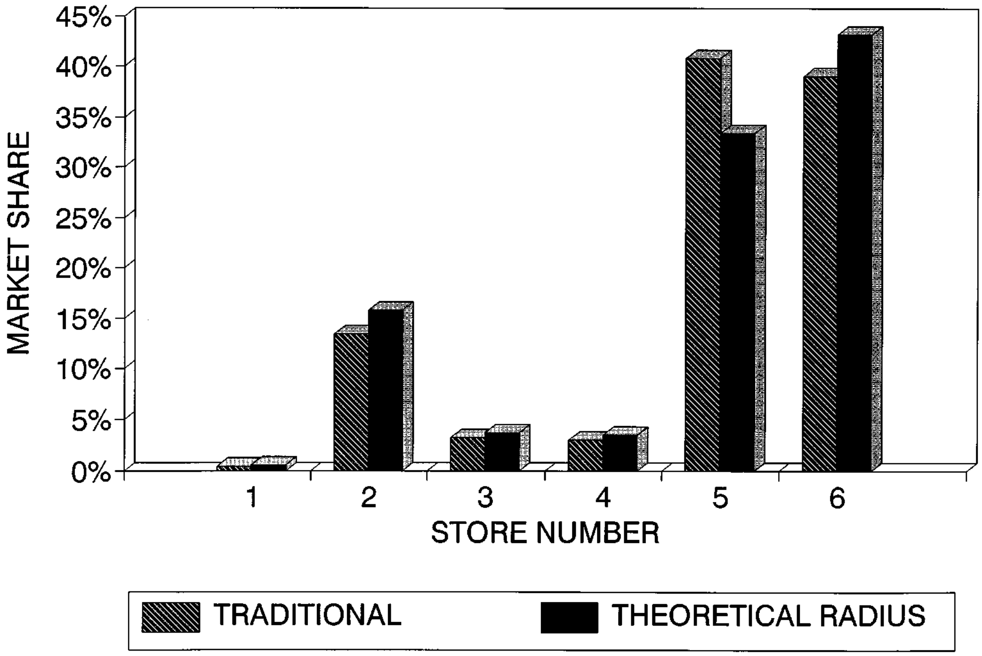


Exhibit 6
Market Share in Tract #090034974 for Existing Six Stores

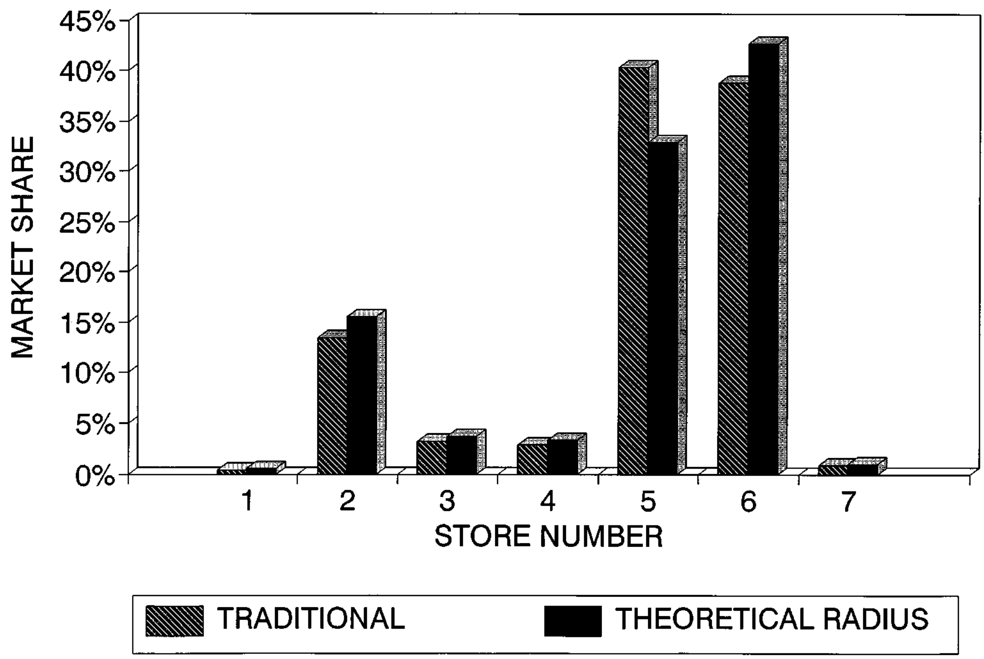


Next, we introduce a new competitor and re-estimate market shares. When this analysis is repeated to consider a potential new competitor the estimated market share differs substantially from the first analysis. The new competitor is capturing 13.6% of the market. Similar to the results discussed earlier, there appears to be little difference between the overall market shares arrived at from using traditional distance measures and those arrived at by employing the theoretical radius. However, there is once again evidence of notable differences in the market share allocations estimated for the Census tracts that actually made use of the theoretical distance measures. Exhibit 7 shows the initial market share estimates for stores 5 and 6 are 40.4% and 38.8%, respectively, for tract 090034974. The theoretical radius provides market share estimates for stores 5 and 6 equal to 33.0% and 42.7%, respectively, for the same Census tract (# 090034974). Exhibit 8 presents differences in market share estimates, traditional minus theoretical radius, for Census tract 090035018.

Evaluation of Results

Anytime there is a retail site very close to the center of a Census tract, (or other geographic area under study), traditional market share estimates will not likely be based on accurate measures of the actual travel distance faced by consumers. Although the details of these findings are specific to this study, differences in estimated market shares should be expected in other markets. The theoretical radius, easily obtained from a GIS, can provide better estimates of the actual distance faced by consumers.

Exhibit 7
Market Share in Tract #090034974 for Seven Stores
(7th store is a new competitor)



It is of course possible that the theoretical radius also does not provide an accurate estimate of the actual travel distance confronted by consumers. This may occur when the geographic area under study is grossly different from a circle (see the discussion in the second section).

One possible criterion for establishing the soundness of the theoretical radius is to make use of another measurement easily available from a GIS. This measure is the perimeter of any geographic area such as a Census tract. As we previously discussed in the second section, a ratio can be calculated comparing the actual perimeter to a theoretical perimeter. The theoretical perimeter can easily be calculated from the theoretical radius; $P_c = 2\beta r^*$, where r^* is the theoretical radius. The extent to which the ratio diverges from one provides an estimate of how much a subject area differs from a circle. If the ratio suggests that a circle is not an adequate approximation for the subject area, other measures should be used instead of the theoretical radius.[16] However, if the ratio suggests that a circle is a reasonable approximation to the subject area, the theoretical radius provides an easily available alternative to the traditionally used distance measure employed in gravity models.

For example, we feel that overall the Census tracts in the trade area under study do not grossly diverge from a circle. However, there are a few that may be vastly different

Exhibit 8
Market Share Differences in Tract #090034974 for Seven Stores
Differences Are Traditional Market Share Estimates Less Estimates Derived by Employing the Theoretical Radius

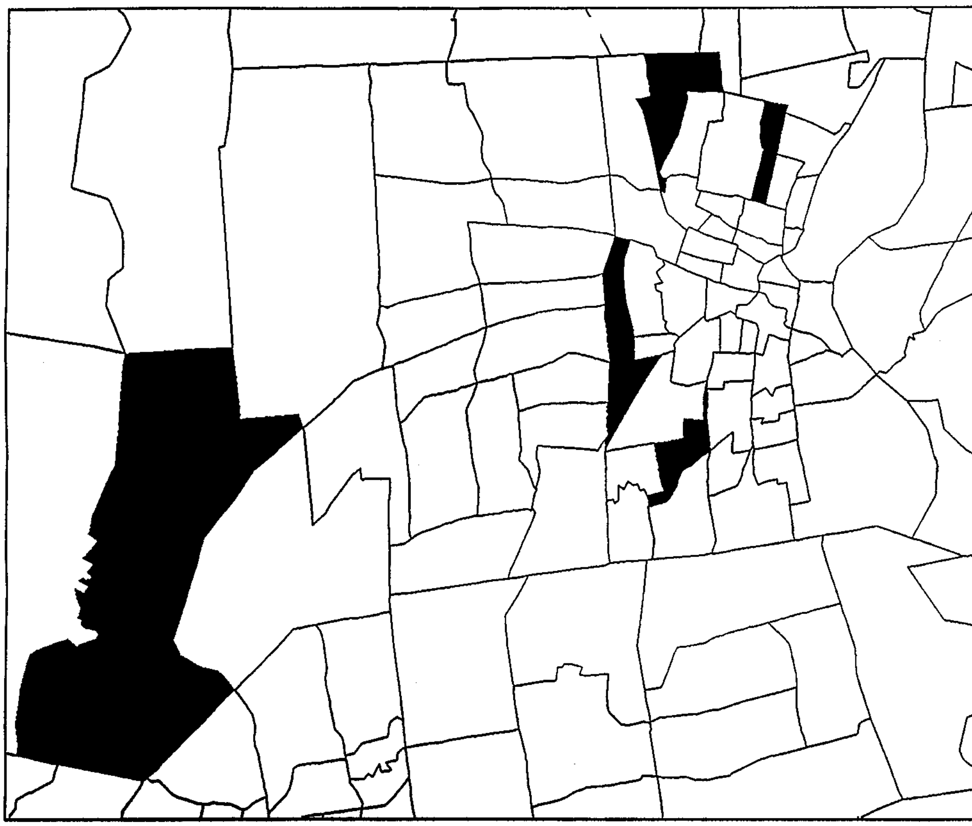
from a circle, and hence for those cases it may be necessary to obtain an alternative to the theoretical distance measure. We calculate perimeter ratios for each Census tract as describe above and screen out the 10% of the Census tracts most unlike a circle based on the ratio of perimeters.[17] Exhibit 9 highlights the Census tracts identified by this criterion. A GIS allows for easy implementation of this type of screen and appears to have accurately chosen those tracts that are grossly unlike a circle.

Alternate Distance Measures: The Need for Software Development

The theoretical radius was not employed in any of the highlighted Census tracts in Exhibit 9 since no relevant retail site is located there. If a store was located near the centroid of tract 09003460202 (the left-most highlighted tract in Exhibit 9), what would be an appropriate distance measure for the gravity analysis?

Three alternative measures of linkages are presented at the end of the second section. The second of these alternatives (using average travel distances along the road network), is appealing for several reasons. Without address-specific data, the road network provides the best information as to the location of homes because most homes will tend to be located along roads. (However, there exists the possibility of industrial areas or other areas without homes along the roads.) Assuming that the population of

Exhibit 9
Census Tracts in Trade Area Identified to Grossly Diverge from a Circle Based on Ratio of Actual Perimeter to Theoretical Perimeter

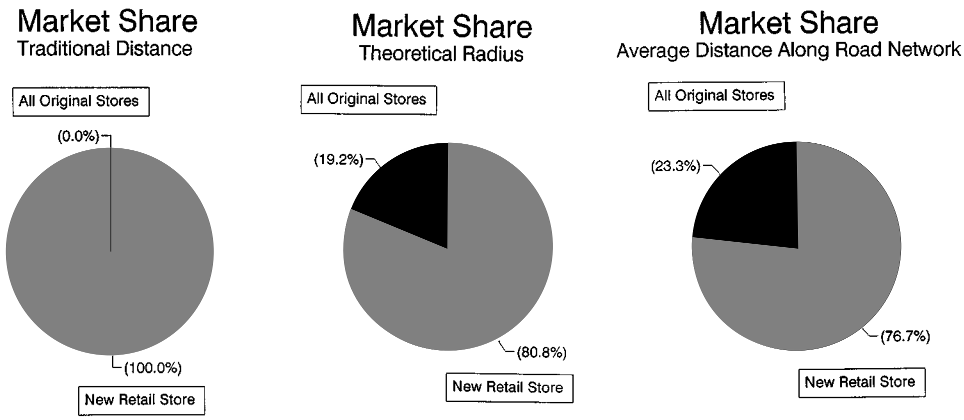


consumers is uniformly distributed along the road network should provide more accurate distance measures than assuming a uniform distribution over the entire geographic area. The latter assumption will weigh areas such as lakes or parkland the same as neighborhoods, whereas the former assumption will only consider areas where roads exist.[18]

With varying degrees of effort, most GIS software can be used to provide this information. Furthermore, given the high paced advances in GIS software, greater development of vertical applications seems inevitable. The arguments in this article point to a need for GIS software that can easily calculate average travel distances along roads within a polygon.

The theoretical radius for tract 09003460202 is 1.559 miles. The perimeter ratio analysis suggests that using two-thirds of this distance (1.039 miles) may not reflect the actual travel distance consumers would face if a retail site was located near the

Exhibit 10
Market Share Estimates for a New Retail Site Located Near the Centroid of Tract #09003460202 Employing Three Alternative Distance Measurements



centroid of this tract. Given the assumption of a uniform distribution of consumers along the road network, a GIS provides an average travel distance along the road network equal to 1.176 miles. This calculation may require some programming in a GIS.

Exhibit 10 shows the market share estimates for a new retail site near the centroid of tract 09003460202. Distance to the centroid (the traditionally used distance measure) provides estimates of 0% market share for all competitors and 100% market share for the new fictional retail site near the centroid of the Census tract.[19] The theoretical radius gives an estimate of 80.8% market share for the retail site near the centroid of this tract. Finally, the average travel distance along the road network provides an estimated market share of 76.7% for the retail site near the centroid of the tract. The last result provides the best estimate of market share given the commonly available grainy data and other assumptions made herein. It is interesting to note that even in the case where the geographic area grossly differs from a circle (like tract #09003460202), the theoretical distance provides a reasonable alternative to the traditionally used distance measure.

Conclusion

Much of the information used in real estate market analysis is currently available only at levels of spatial aggregation above the address level. This poses a problem for research on urban economics and real estate: How can linkages be accurately measured? Typically, spatially aggregated data are associated with the centroid of the geographic area, but distance to the centroid may understate the actual distance faced by consumers.

We explain how to calculate theoretical distances that can be substituted for distance to the centroid. The theoretical radius (the radius of a circle with the same area as the Census tract) is relatively easy to implement with existing GIS software. The theoretical radius performs well when the actual shape of the geographic area is a square, a rectangle, or irregular in shape, provided that the geographic area is not shaped too much like an elongated rectangle (*e.g.*, a rectangle with one side three or more times the length of the other side).

Our theoretical and empirical results demonstrate that the ratio of the perimeter of the geographic area to the theoretical perimeter is useful for determining when the theoretical radius is a poor approximation to actual travel distance.[20] In this case (*i.e.*, when the shape of the area is sufficiently unlike a circle) several alternative measures of average travel distance are explored:

1. If the geographical area is similar to a square or a rectangle, then the formulas derived in the technical appendix can be used.
2. GIS software may be used to calculate average travel distance along roads.
3. Numerical methods can be used to approximate average travel distance.

With most existing GIS software, method 1 will be easier to implement than method 2, which will be easier than method 3.

We use a GIS with actual data to illustrate how to handle spatially aggregated data for a gravity model. If traditional distance measures are employed, the closer a retail site is to the centroid of a geographic area being studied, the higher will be its estimated market share. This will occur regardless of the real distribution of consumers across the geographic area. The theoretical distance will provide a more realistic measure of the actual distance faced by consumers. Also, it is possible to use GIS software to calculate average travel distance along roads; the amount of programming required depends on the GIS software. We advocate development of GIS software that facilitates future analysis of commonly available spatially aggregated data.

Appendix

Derivation of Average Travel Distance Equations

The purpose of this Appendix is to derive equations for average travel distance on a circle, a square and a rectangle. The assumption throughout is that there is one person per square unit of distance. The constant one is arbitrary and has no effect on average travel distance.

A person at any point on the circumference of a circle travels distance r to reach the center, so the total travel distance, TT, for persons on the circumference is proportional to:

$$TT = (\pi D)r = 2\pi r^2, \tag{A1}$$

where TT = total travel distance for all people living on the circumference of a circle with radius r.

To find the total travel distance for all the people within the circle (as well as on its circumference), we need "sweep up" all the people for circles from radius 0 to radius r. This integral is:

$$TT_c = \left(\frac{2}{3}\right) \pi r^3, \tag{A2}$$

where TT_c = total travel distance for all people living in the circle of radius r. Since there are πr^2 people living in the circle, the average travel distance is:

$$T_c = \frac{2}{3} r = .6667r. \tag{A3}$$

Formula for the Square

The general outline of the logic for travel distance on a square is the same as for a circle. We begin by drawing a perpendicular from the center of the square to one side; the travel distance for the person living at the point where the perpendicular intersects the side is $x/2$. It is convenient to rescale by letting $X_1 = x/2$. Then the average travel distance for the people living on this half side is given by:

$$\int D = \int_{y=0}^{y=X_1} D(X_1, y)\, dy = \int_0^{X_1} (X_1^2 + y^2)^{.5}\, dy, \tag{A4}$$

where y is the vertical coordinate that fixes the location of the person on the side of the square. We can simplify this equation with:

$$\int D = X_1 \int_0^{X_1} \left(1 + \left(\frac{y}{X_1}\right)^2\right)^{.5} dy. \tag{A5}$$

To evaluate the average travel distance of a person living on a unit surface, we have:

$$\int D = X_1 \int_0^1 (1 + f^2)^{.5}\, df = X_1 S. \tag{A6}$$

This equation has a closed form solution given by formula 156, page 343 of the CRC Tables (Beyer, 1987). To find the total travel distance for all people living on the half side, we multiply by X_1:

$$TT = X_2 S = 1.1478 X_1^2. \tag{A7}$$

Now, we "sweep up" all the people from an X_0 to X to obtain total travel distance on a square with side X_1:

$$TT_s = 1.1478\, X^3/3 = .3826\, X^3. \tag{A8}$$

Since X_1 is a half side, there are eight times this travel distance on the big square. But when we substitute for X_1, we obtain $T_s = .3826x^3$ for the big square. Since there are x^2 people living on the square, the average travel distance is:

$$T_s = .3826x \tag{A9}$$

The logic of this integral was checked with numerical integration using Mathematica. The two results agreed to two decimal places. Similar calculations using numerical summations in Gauss agree.

Average Travel Distance on a Rectangle

We begin by dividing the rectangle into quadrants; a perpendicular is dropped from the center to one of the sides. Then a perpendicular is drawn to the other side to form a rectangle with area of one quarter of the big rectangle. These two half sides are labeled X_1 and X_2. There are $4X_1X_2$ people living in the large rectangle.

Following the same logic as for the square, it is possible to derive a closed form solution for the rectangle, but this solution is exceedingly cumbersome to work with because it depends on $z = X_2/X_1$. It turns out to be much easier to tabulate formulas for average travel distance on the big rectangle as a function of z:

z	T_R
1	$.3826X_1$
2	$.5932X_1$
3	$.8231X_1$
4	$1.0608X_1$
5	$1.3023X_1$
.	
.	
.	
.	
10	$2.53192X_1$

where X_1 is the short side of the rectangle. Note that the increments to the constant multiplied by X_1 asymptote to about .246 for each unit of z.

Notes

[1] For an introduction to GIS, see Star and Estes (1990) and Huxhold (1991). Thrall and Marks (1993) and Marks and Thrall (1994) provide a discussion of GIS in real estate research.
[2] We use the term "aggregation" to mean spatial groups of economic agents such as households or employees. Arbia (1989), on the other hand, uses the term for the shape of an area. Thus, he deals with the problem of *how* to group small areal units into a given number of larger areas.
[3] For example, Thorson (1994) uses distance measurements to the central business district (CBD) and to nearest town in a study of the effects of zoning changes on land markets. In their study of housing price gradients in a multinodal urban area, Waddell, Berry and Hoch (1993)

use a GIS to calculate the straight line distance from the centroid of neighborhoods to each of the major employment centers, as well as to the nearest freeways and minor and major retail sites. Do, Wilbur and Short (1994) use straight-line distance measurements from residential properties and the nearest church. Bible and Hsieh (1993) use straight-line distance measurements to schools and shopping centers in their hedonic model of apartment rents.

[4] For example, Benjamin, Jud and Winkler (1995) use data at the state level and national level in their study of shopping center investment, and Ambrose and Springer (1993) use data at the country level and MSA level. The potential problems are not constrained to domestic data. Cheung, Tsang and Mak (1995) use spatially aggregated data for five categories of property located in four districts of Hong Kong.

[5] The centroid is typically the point that minimizes travel distance from the boundaries of the area. In the past, this location was painstakingly found manually.

[6] Arbia (1989) summarizes a large literature dealing with issues of spatial scale and aggregation. The appropriate levels of scale and aggregation are the subject of this literature, whereas we take these levels as given by the data. Instead, we focus on measurement of linkages to economics agents within an area used to aggregate these agents.

[7] Rodriguez, Sirmans and Marks (1995) explain how a GIS can be used to calculate the shortest path along the road network and that this distance measurement more accurately reflects actual travel distances.

[8] Address level data solves this problem, but many important types of data are simply unavailable at the address level. Furthermore, it is expensive to maintain current address data.

[9] The problem here is similar to the measurement of within-group connectedness (Arbia, 1989: 36–7). Also, Can and Megbolugbe (1996) use a measure of compactness, which is related to our concept of theoretical radius. However, unlike this literature, we address linkages to economic agents within a geographical area.

[10] The two-thirds constant assumes a uniformly distributed population, a reasonable assumption in the absence of address data. Alternate assumptions will be examined later.

[11] This is done with two Gauss programs designed to find the centroid and then the average travel distance for irregular geometrical shapes.

[12] A similar conclusion holds for triangles. The circle is a good approximation to an equilateral triangle. But, as an isosceles triangle becomes more distorted away from the equilateral triangle, the theoretical radius can become a poor approximation.

[13] In the remainder of this paper, P_i = perimeter of shape i (s = square, c = circle and R = rectangle).

[14] Wilson (1970) provides an alternative where the distance impedance parameter enters in a negative exponential form, $e^{-\lambda Dij}$. In cases where the traditionally used distance between a retail site and a neighborhood approaches zero, distance drops out of the analysis. The attractiveness measure will not reflect the actual distance faced by consumers, but will only be a function of retail site characteristics.

[15] We use Equation (11) with $\beta = 1$ and $\lambda = 2$.

[16] As discussed later, a GIS could be used to establish the average travel distance along the road network faced by consumers.

[17] The analyst may establish other screens besides accepting the top 90% of the geographic areas under study. Selection of the appropriate screen should be done in light of the sensitivity of the decision to error.

[18] In a developing area it is important to use maps within a GIS that contain the latest road segments so that new neighborhoods are not ignored in the analysis.

[19] The reason for these extreme results is that the new fictional retail site is located almost at the centroid of the tract, making the attractiveness of the retail site approach infinity.

[20] This ratio is easy to calculate with most GIS software.

References

Ambrose, B. W. and T. M. Springer, Spatial Variation of Nonmetropolitan Industrial Location, *Journal of Real Estate Finance and Economics*, 1993, 7, 17–27.

Arbia, G., Spatial Data Configuration in Statistical Analysis of Regional Economic and Related Problems, London: Kluwer Academic Publishers, 1989.

Benjamin, J. D., G. D. Jud and D. T. Winkler, An Analysis of Shopping Center Investment, *Journal of Real Estate Finance and Economics*, 1995, 10, 161–68.

Bible, D. S. and C. Hsieh, Applications of Geographic Information Systems for the Analysis of Apartment Rents, Draft paper, 1993.

Can, A. and I. Megbolugbe, The Geography of Undeserved Mortgage Markets, Paper prepared for the mid-year meeting of the American Real Estate and Urban Economics Association, May 1996.

Cheung, Y., S. Tsang and S. Mak, The Causal Relationship Between Residential Property Prices and Rentals in Hong Kong: 1982-1991, *Journal of Real Estate Finance and Economics*, 1995, 10, 23–35.

Do, Q., R. W. Wilbur and J. L. Short, An Empirical Examination of the Externalities of Neighborhood Churches on Housing Values, *Journal of Real Estate Finance and Economics*, 1994, 9, 127–36.

Haynes, K. E. and A. S. Fotheringham, *Gravity and Spatial Interaction Models,* London: SAGE, 1984.

Huff, D. L., A Probabilistic Analysis of Shopping Center Areas, *Land Economics*, 1963, 39, 81–90.

Huxhold, W. E., *An Introduction to Urban Geographic Information Systems,* Oxford and New York: Oxford University Press, 1991.

Marks, A. P., C. Stanley and G. Thrall, Criteria and Definitions for the Evaluation of Geographic Information System Software for Real Estate Analysis, *Journal of Real Estate Literature*, 1994, 2, 227–41.

Rodriguez, M., C. F. Sirmans and A. P. Marks, Using Geographic Information Systems to Improve Real Estate Analysis, *Journal of Real Estate Research*, 1995, 10, 163–73.

Star, J. and J. Estes, *Geographic Information Systems: An Introduction*, Englewood Cliffs, NJ: Prentice-Hall, 1990.

Thorson, J. A., Zoning Policy Changes and the Urban Fringe Land Market, *Journal of the American Real Estate and Urban Economics Association*, 1994, 22, 527–38.

Thrall, G. and A. P. Marks, Functional Requirements of a Geographic Information System for Performing Real Estate Research and Analysis, *Journal of Real Estate Literature*, 1993, 1, 49–61.

Waddell, P., B. J. Berry and I. Hoch, Residential Property Values in a Multinodal Urban Area: New Evidence on the Implicit Price of Location, *Journal of Real Estate Finance and Economics*, 1993, 7, 117–41.

Wilson, A. G., *Entropy in Urban and Regional Modeling,* London: Pion Limited, 1970.

Research support was provided by the Charles Tandy American Enterprise Center of the M. J. Neeley School of Business, Texas Christian University and the Center for Real Estate and Urban Economic Studies at the University of Connecticut. Dr. Clapp is grateful to Joe Rando for discussion of the ideas in this article.