ELSEVIER

Available online at www.sciencedirect.com

SciVerse ScienceDirect

Procedia Environmental Sciences 7 (2011) 287–292

Procedia Environmental Sciences

1[st] Conference on Spatial Statistics 2011

Spatial Statistics Methods in Retail Location Research: A Case Study of Ankara, Turkey

Burcu H. Ozuduru[a], Cigdem Varol[b], a*

[a] Instructor, Gazi University, Faculty of Architecture, Department of City and Regional Planning, Ankara, 06570, Turkey
[b] Assistant Professor, Gazi University, Faculty of Architecture, Department of City and Regional Planning, Ankara, 06570, Turkey

Abstract

This study, using two methods of spatial data assessment: 1) spatial autoregression (SAR) models (Cliff and Ord, 1981) and 2) geographically weighted regression (GWR) (Fotheringham et al., 1998), identifies the relationship between shopping centers attributes and trade area (TA) characteristics in Ankara, Turkey. Ankara has the highest level of shopping center gross leasable area per capita in Turkey, for this reason, is a unique case study. The two models provide information on distinctive characteristics of shopping center locations. The first one depicts the *global* relationship between shopping center supply, assessed by total gross leasable area in each district, and demand, assessed by demographic and socioeconomic characteristics of TAs, such as population, income, homeownership while accounting for the spatial dependence across the TAs. The second one, on the other hand, assesses *local* demand supply relations at the district level. These two models do not substitute but complement each other. SAR model results show that there is a positive relationship between shopping center supply and median age, and a negative relationship between supply and household size. These are expected results in compliance with the literature findings. The level of homeownership variable, however, illustrates a distinctive picture, unique to Ankara, in that there is a negative relationship between homeownership and shopping center supply. The GWR results show that it is easier to explain the level of variations in selected parameters in the suburbs than in inner city neighbourhoods, therefore, as expected, in car dependent suburbs stronger relations with shopping center locations are identified than mixed-use inner city neighbourhoods. The results are essential for identifying the spatial network of retail outlets in a city or region which guides urban developers, investors and public policy decision makers in site selection.

Selection and peer-review under responsibility of Spatial Statistics 2011

Keywords: Retail Location; Shopping Centers; Spatial Auto Regression; Geographically Weighted Regression

1. Introduction

Deciding where to locate a shopping center is a vital part of a retail business strategy, in that each center becomes a part of an existing network, and is located in relation to the market characteristics and other shopping center locations in a trade area. These trade areas are often spatially dependent, due to the

* Corresponding author. Burcu H. Ozuduru. Tel.: +90-312-582-3701; fax: +90-312-582-3661.
E-mail address: bozuduru@gazi.edu.tr.

Selection and peer-review under responsibility of Spatial Statistics 2011
doi:10.1016/j.proenv.2011.07.050

fact that customers residing in one location may shop at another, indicating that shopping center sales at one location are dependent on the consumer characteristics of another. For this reason, retail site selection process becomes a complex process that needs special focus. The spatial dependency between trade areas is assessed by spatial autocorrelation, which is measured by a number of statistics having slightly different formulations but they depend on some definition of spatial weighting which attempts to quantify the often subjective concept of proximity, in other words, what is meant by 'nearby'? [1]. It is important to identify this in order to depict true relations across observations in space. Initially, spatial autocorrelation models were used to depict the causes of deaths in the UK [2]. However, this task has spread over to other topics, such as assessment of the spatial distribution of house prices [3], characteristics of housing sub-markets and their sale prices [4] and neighborhood location characteristics and dwelling sale prices [5]. In retail real estate literature, such research has been carried out only for retail real estate valuations when assessing the relationship between the characteristics of a store and total rent [6] but their application in retail location analysis has been limited. In this paper, we analyze the spatial distribution of shopping centers by assessing the relationship between retail supply, proxied by total gross leasable area of shopping centers per capita (TGLAPCAP), and trade area characteristics, proxied by demographic and socioeconomic attributes of the residents in each district by using spatial auto regression (SAR) and geographically weighted regression (GWR) models. This assessment intuitively embraces spatial dependence, thus is an intriguing subject of spatial autocorrelation. The two models provide information on global and local variations of the trade area relations in specified geographical units, districts, with reference to specific spatial weighing schemes. This unique approach offers an analysis of trade area relations in assessment of retail supply levels in each area, thus offers an evaluation of the spatial network of shopping centers in a larger geographical unit, in this case, in Ankara, which has the highest level of total gross leasable area per capita in Turkey and the amount of shopping center space supply has increased rapidly since 2007. Using the two models, analysts can understand and specify: 1) the level of retail supply in a district in relation to the districts' trade area characteristics, whether the district is over- or under- supplied, 2) significant local trade area attributes substantial in retail site selection, 3) spatial network of retail outlets in a city or region. In other words, these models, apart from traditional location selection models, offer a geographic framework for intra- and inter-relational aspects of retail real estate dynamics.

2. Literature Review

A fundamental property of spatially located data is that the set of values, x_i, are likely to be related over space [7] The main assumption of SAR models is that each geographical unit has to be considered in relation to its surrounding geographical units, and for this reason, observations in a geographical unit are also dependent on observations in these units [8]. Such relations indicate that variables and error terms are dependent. The spatial autocorrelation, which can be specified both in time and space reveals the functional dependencies of areas. Retailing facilities can either supplement or complement each other depending on their facility type, thus either are clustered or evenly distributed over space. The trade area attributes of a geographical unit is not always distinctive for a retail outlet's success. It is true that the customers shop at the closest outlet for convenience goods; however, for shopping goods, which are more unique and assorted, the situation is more complex. When shopping goods are of concern, customers do not shop at the closest retail outlet but shop at distant locations and travel longer distances to find a particular shopping good that they need. Therefore, the customer of a shopping center is not always a resident of the closest trade area but is a resident of another shopping center's trade area. This mutual interaction points to the concept of *geographical dependence* of shopping center trade areas. The concept was first raised by Russell [9], who related consumer out-shopping to geographical interdependency in urbanized areas for retail establishments, and points to the ambiguity of TA model results that ignore the phenomenon. Similar points have been made by several researchers [10]. Moran's Test is particularly useful for the assessment of this dependence. The Moran's I statistic is formulated in terms of regression residuals and spatial connectivity weights [11]. The observations in each unit are compared under the null hypothesis, H_0, of no spatial auto-correlation. SAR models are used when the model's error term is heteroskedastic and is related to a certain variable that is not independently distributed over space; thus, the model variables and error terms are spatially correlated [12]. The units are negatively spatially correlated if the income variable in each unit is of different levels. When this happens, the ordinary least squares (OLS) assumption of independently and identically distributed residuals is violated. The spatial auto-correlation is accounted for through the construction of the SAR model:

$$Y_i = X_i\beta + \varepsilon_i \quad (1)$$

$$\varepsilon_i = \rho W_i \varepsilon_i + u_i, \quad (2)$$

where i is an element of all units from i to j; X_i is the (n_i x k) matrix of explanatory variables associated with unit i; W_i is the (i x j) spatial weight matrix of geographic units. The disturbance terms depend on u_i in an autoregressive way, being $u_i \sim N(0, \sigma^2 I_i)$, where I_i is the covariance vector of Y_i. In this model ρ is the parameter that captures the magnitude of the spatial auto-correlation; and β is the vector of unknown parameters associated with the exogenous variables X_i. GWR models are based on the traditional regression framework, and incorporate local spatial relationships into this framework by allowing local rather than global parameters to be estimated. The model is rewritten as:

$$y_i = \beta_0(u_i, v_i) + \sum_k \beta_k(u_i, v_i) x_{ik} + \varepsilon_i \quad (3)$$

where (u_i, v_i) denotes the coordinates of the i[th] point in space and $\beta_k\ (u_i, v_i)$ is a realisation of the continuous function $\beta_k\ (u, v)$ at point i. Thus, parameter values and measurements on a continuous surface are taken at certain points and they denote the spatial variability of that surface [13]. GWR equation recognizes these spatial variations in relationships and provides a way that they can be measured [14]. The coefficients are not random but are deterministic functions of some other variables, which is location in space. The observation is weighted in relation to this location and on a continuous surface varies with location. GWR offers an opportunity to distinguish the non-stationary process of a model that is reflected through the error terms in a model.

3. Data

Initially, a continuous surface indicating shopping center supply, total gross leasable area, is constructed. The supply is distributed across space in relation to primary, secondary and tertiary trade areas, with 100%, 50%, and 25% to primary, secondary and tertiary trade areas, which cover the districts within 0-15, 15-25, and 25-35 minutes driving time. The summation of retail supply in each district is divided by total population, providing the dependent variable, TGLAPCAP. The demand variables, such as population, age distribution, employment status, education, etc. are assembled from the Turkish Statistical Institute database pertaining to the year 2000.

4. Methodology and Model Specifications

The spatial dependence across districts in Ankara in terms of TGLAPCAP is assessed by Moran's Test. The index, I, shows that there is a certain level of spatial dependence across these districts. Next, several SAR models are constructed. Prior to the model construction, the correlations across all variable pairs are carefully analyzed. In particular, the correlation between variables resembling similar qualifications is analyzed. For example, education levels and occupation sector have high correlations than more distinctive variables, such as age and household size because the former variable pair resembles income levels whereas the latter resembles two distinctive qualifications, which are not correlated. The models are formulated to avoid including such variables simultaneously. SAR models are constructed with the dependent variable, TGLAPCAP in each district, and independent variables. Demographic variables consist of total population, age and age groups and socioeconomic variables education, occupation, employment status, household size, tenure type. In order to obtain a comparison in relation to population, the shares of age, education, occupation have been calculated by dividing each value by the total population in each district. Next, the logarithms of all variables are computed. The system of equations at the district level is formulated as follows:

$$TGLAPCAP_i = \boldsymbol{P_i \alpha_i} + \varepsilon_i \quad (4)$$

$$\varepsilon_i = \rho \mathbf{W}_i \varepsilon_i + \mathbf{u_i}, \quad (5)$$

where $\boldsymbol{P_i}$ represents the logarithm of each characteristic of the TAs in selected ZCU, i; and α_i is the parameter estimate. The spatial weights matrix, $\mathbf{W}_i$, is retrieved by constructing the lattice structure to assess the contiguity of the districts. The shared polygon boundaries of each pair of districts are assessed using the R programming language, by which a neighbors list is obtained. This list is in the form of a row standardized contiguity matrix that displays neighboring district weights and sum up to 1. Selected significant variables are used to construct the GWR models. GWR models constructed using the GWR 3.0 software. Results are imported to ArcGIS and local R^2 and retail supply estimate differences are displayed on maps.

5. Empirical Results

5.1. Spatial auto regression model results

The Moran's Test shows a high level of spatial dependence ($I = 0.3664). SAR models are constructed by using the free software R [15]. The model types are implemented in R in the library 'spdep' [16]. The SPSS software is used to run the preliminary analysis of model variables, such as factor and bivariate correlation analyses. These two analyses have been helpful to identify the significant variables and possible multicollinearity across variables. Selected variable results are shown in Tables 1 and 2. The goal is to understand how each variable relates to the dependent variable, TGLAPCAP. It is observed that the homeownership, household size, ratio of employed population to unemployed have negative, and the share of population with high school degree, average age, share of population excluded from the labor force, ratio of women to men with college education variables have negative relations with TGLAPCAP. The negative relationship between TGLAPCAP and homeownership in the districts of Ankara is unique due to the fact that homeownership is higher in poorer districts in Ankara because the squatter residents are regarded as homeowners. Both significance levels and model indicators, Akaike Information Criterion (AIC) is considered to assess the reliability of the models. A reduction in the AIC from the models suggests that the model is better in accounting for differences in degrees of freedom (Model #1 generates better results than Model #2). Model #2 variables are used in formulation of GWR models.

Table 1. SAR Model Results #1

Independent variables	**Estimate**	**Standard error**	**Z-value**	**P-value**
Intercept	-4.212	1.300	-3.240	0.001
Share of population with high school degrees	1.006	0.266	3.784	0.000
Average age	1.806	0.402	4.493	7.009e-06
Share of homeowner households	-0.528	0.143	-3.684	0.000
N= 351 AIC=528.99 (AIC for lm=527.01)				

Table 2. SAR Model Results #2

Independent variables	**Estimate**	**Standard error**	**Z-value**	**P-value**
Intercept	6.094	0.299	20.408	<2.2e-16
Share of population excluded from the labor force	0.312	0.124	2.525	0.012
Ratio of employed population to unemployed	-1.050	0.056	-18.685	<2.2e-16
Household size	-1.449	0.290	-5.005	5.58e-07
Ratio of women to men with college education	0.217	0.110	1.976	0.048
N=376 AIC=390.34 (AIC for lm=389.78)				

5.2. Geographically weighted regression model results

Table 3. Global regression parameters

Global GWR Results	**Values**		
Akaike information criterion	394.124		
Coefficient of determination	0.488		
Adjusted r-square	0.481		
Parameter	**Estimate**	**Standard error**	**T-value**
Intercept	6.221	0.282	22.092
Share of population excluded from the labor force	0.329	0.221	1.487
Ratio of employed population to unemployed	-1.046	0.057	-18.391
Household size	-1.488	0.296	-5.021
Ratio of women to men with college education	0.227	0.113	2.016

The first section of the GWR output (Table 3) contains parameter estimates and standard errors from a global model fitted to data. These global results suggest that the share of population excluded from the labor force and ratio of women to men with college education variables are positively related to TGLAPCAP. The Ratio of employed population to unemployed and household size variables are negatively related to TGLAPCAP. The model replicates the data reasonably well (48% of the variance in TGLAPCAP is explained by the model). There are some factors not captured by the global model. The first section of GWR is a diagnostic for the GWR estimation followed by similar information to the corresponding panel for the global model (Table 4). The second section is constructed from a comparison of the predicted values from different models at each regression point and the observed values [17]. The coefficient has increased from 0.488 to 0.633. The reduction in the AIC from the global model suggests that the local model is better even accounting for differences in the degrees of freedom [18]. The main output from GWR is a set of localised parameter estimates and associated diagnostics, which lend themselves to being mapped [19]. The local r-square statistic in the GWR model shows that the model variables are explanatory in particular in the inner city districts.

Table 4. GWR Estimation

Fitting GWR		**Diagnostic Information**	
Number of observations	376	Residual sum of squares	41.073
Number of independent variables	5 (Intercept is variable 1)	Akaike Information Criterion	283.280
Bandwidth (in data units)	5195.889	Coefficient of Determination	0.654
Number of locations to fit model	376	Adjusted r-square	0.633

In the final step, the final SAR model is computed for each district by plugging in the valid district values of specific variables. The results are subtracted from the existing total shopping center gross leasable area per capita supply (Fig. 1). The districts with light colors are undersupplied; the difference between the existing level and model result is low (less than 1.75). The darker color districts satisfy the demand of the existing residents when their specific characteristics (included in the SAR model) are considered. The range is between 1.76 and 3.00. Finally, the darkest color districts are over supplied in relation to their characteristics. The map shows that, most of the inner city districts are supplied at the expected levels. However, there are some districts, in particular, on the northern parts of Ankara are under supplied. When the demographic and socio economic characteristics of these districts are analyzed, it is observed that these districts are relatively lower income districts. On the west in recent suburbs of Ankara, there is a potential for shopping center investment.

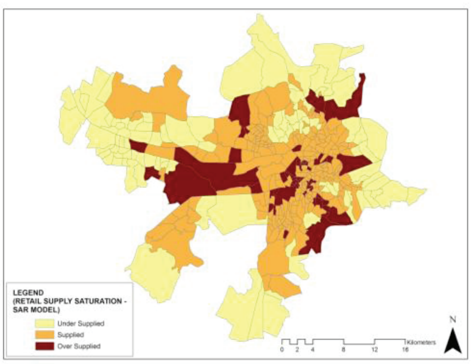

Fig. 1. The Difference between Existing and SAR Results

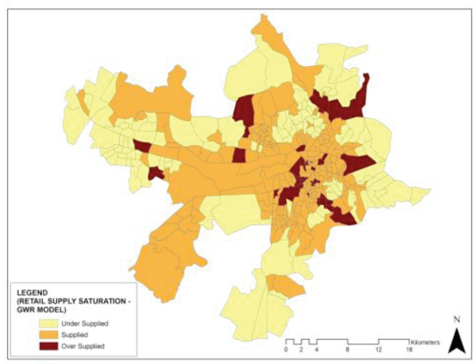

Fig. 2. The Difference between Existing and Local GWR Results

The local GWR results are significantly different than SAR results (Fig. 2). These are calculated by subtracting the local parameter estimates and values plugged in the model from the existing TGLAPCAP. The results show the number of under supplied districts are more and over supplied is less. In particular, the over supplied districts on the western part disappear and more under supplied districts appear on the southern part, pointing to potential investments. The global GWR model results are similar to SAR results pointing to the consistency across global models.

6. Conclusion and Discussion

The case study shows that SAR and GWR models are suitable for assessing the relationship between trade area characteristics and shopping center attributes, which accounts for the significant spatial dependence identified across the districts. A reduction in AIC and increase in R^2 points to this suitability. The GWR results show that selected variables, which are share of population excluded from the labor force, ratio of employed population to unemployed, household size, ratio of women to men with college education explain the changes in inner city districts better than the outer city districts. This result is also supported with the local R^2 values in inner city districts. In addition, the GWR results show a significant spatial variance of *individual variable variations* between inner city districts and outer city districts on the West and North. More specifically, it shows that the supply in most inner city districts are supplied or over-supplied and there is an investment potential on the outer city districts in particular on the West and North. Models formulated in this study can be used for location selection of shopping centers, identification of crucial variables in retail supply variation at particular geographical units and levels, finally, in specification of the overall spatial network of shopping centers in a city or region. The model results are useful to investors, developers, policy makers and local and central authorities effective in shopping center development processes.

Acknowledgement

We thank the Scientific and Technological Research Council of Turkey (TUBITAK) for the financial and institutional support for conducting the research of this study.

References

[1] Fotheringham AS, Charlton ME, Brunsdon C. *Geographically weighted regression: The analysis of spatially varying relationships*. Hoboken, New Jersey: John Wiley and Sons Ltd; 2002.

[2] Cliff AD, Ord JK. *Spatial processes: Models and applications*. London: Pion Ltd; 1981.

[3] Fotheringham AS, Charlton ME, Brunsdon C. *Geographically weighted regression: The analysis of spatially varying relationships*. Hoboken, New Jersey: John Wiley and Sons Ltd; 2002.

[4] Bourassa SC, Cantoni E, Hoesli M. Spatial Dependence, Housing Submarkets, and Housing Price Prediction. *Journal of Real Estate Finance and Economics* 2007;**35(2)**: 143-160.

[5] Militino AF, Maria DU, Garcio-Reinaldos L. Alternative Models for Describing Spatial Dependence among Dwelling Selling Prices. *Journal of Real Estate Finance and Economics* 2004;**29(2)**: 193-209.

[6] Carter CC, Haloupek WJ. Spatial Autocorrelation in a Retail Context. *International Real Estate Review,* 2000;**3(1)**: 34-48.

[7] [8] Cliff AD, Ord JK. *Spatial Processes:Models and Applications*. London: Pion Ltd; 1981.

[9] Russell, VK. The Relationship between Income and Retail Sales in Local Areas. *Journal of Marketing* 1957;**21(1)**: 329-332.

[10] Anderson CH, Kaminsky M. The Outshopper Problem: A Group Approach for Small Retailers. *Entrepreneurship Theory and Practice* 1985;**9(3)**: 34-45. Ghosh A, McLafferty SL. *Location Strategies for Retail Firms and Service Firms,* Lexington, MA: D.C. Heath Company 1987. Mejia LC, Benjamin JD. What do we know about the Determinants of Shopping Center Sales? Spatial vs. Non-Spatial Factors. *Journal of Real Estate Literature* 2002;**10(1)**: 3-26. Jarratt DG. Outshopping Behavior: An Explanation of Behavior by Shopper Segment Using Structural Equation Modeling. *The International Review of Retail, Distribution and Consumer Research* 2000; **10(3)**: 287-304. Thompson JR. Characteristics and Behavior of Out-Shopping Consumers. *Journal of Retailing* 1971; **47(1)**: 70-80.

[11] Cliff AD, Ord JK. *Spatial processes: Models and applications*. London: Pion Ltd; 1981.

[12] Kelejian HH, Robinson DP. The Influence of Spatially Correlated Heteroskedasticity on Tests for Spatial Correlation. In: Anselin L, Rey SJ, Florax RJGM, editors. *Advances in spatial econometrics: Methodology, tools and applications*, Berlin Heidelberg: Springer-Verlag; 2004, p. 79-91.

[13] [14] Fotheringham AS, Charlton ME, Brunsdon C. *Geographically weighted regression: The analysis of spatially varying relationships*. Hoboken, New Jersey: John Wiley and Sons Ltd; 2002.

[15] R Development Core Team. R: A Language and Environment for Statistical Computing. R foundation for Statistical Computing, Vienna. 2005. Available at: http//www.R-project.org.

[16] Bivand R. spdep: Spatial Dependence: weighing schemes, statistics and models. R package version 2.11.1. 2006. Available online at http://cran.r-project.org/spdep.html

[17] [18] [19] Fotheringham AS, Charlton ME, Brunsdon C. *Geographically weighted regression: The analysis of spatially varying relationships*. Hoboken, New Jersey: John Wiley and Sons Ltd; 2002.