Spatial Distribution of Retail Sales

R. Kelley Pace
LREC Endowed Chair of Real Estate
2163 CEBA, Department of Finance
E.J. Ourso College of Business Administration
Louisiana State University
Baton Rouge, LA 70803-6308
(225)-578-6256
kelley@pace.am
www.spatial-statistics.com

Ming-Long Lee
Department of Finance
College of Management
National Yunlin University of Science & Technology
123, Section 3, University Road, Touliu, Yunlin, Taiwan 640
(05)-534-2601 EXT. 5338
FAX: (05)-531-2079
leeming@yuntech.edu.tw

To Present at the 10th Annual Conference
of Pacific Rim Real Estate Society

Pace would like to acknowledge support from NSF grant BSC-0136229 as well as from Louisiana State University, the University of Connecticut, and James Stone of Geovue.

Abstract

We examine the distribution of retail sales for a retail chain in the Houston market using a spatial gravity model. Unlike previous empirical studies, our approach models spatial dependencies among both consumers and retailers. The results show that both forms of spatial dependence exert statistically and economically significant impacts on the estimates of parameters in retail gravity models. Contrary to the suggestions of Gautschi (1981) as well as Eppli and Shilling (1996), our results show the importance of the distance parameter in retail gravity models may be greatly understated. Thus, ignoring spatial dependence may lead to overestimation of the deterministic extent of trade areas, and underestimate the importance of good locations.

KEYWORDS: spatial autoregression, spatial statistics, spatial econometrics, gravity model, trade area.

I. Introduction

When considering opening a new store, retail chain store executives take a holistic view of market performance across their entire store network. In particular, they wish to avoid opening a profitable store at the expense of existing stores. To avoid such a loss, executives need the spatial distribution information of customers and competitors to accurately define trade areas for site selection. In addition, managers of individual stores can use the spatial distribution of customers and competitors to promote sales. From an overall market perspective, the technology of forecasting sales can affect the location premia of retail properties.

We apply retail gravitation notions to examine empirically the spatial distribution of retail sales.[1] Over seventy years ago Reilly (1931) published his seminal proposition, known as "the law of retail gravitation." Retail gravity models draw an analogy with Newton's gravitational law to account for human behaviors related to shopping activities. In retail gravity models, various store features such as size attract customers, just as larger astronomical bodies have greater gravitational force. Distance between the customers and the store diminishes this attraction, just as gravity diminishes with distance.

Many studies have examined empirically this concept of retail gravitation.[2] Nevertheless, the literature often differs on the importance of the distance among retailers and consumers (Mejia and Benjamin, 2002). In particular, Gautschi (1981) as well as Eppli and Shilling (1996) suggest that the distance parameter may be significantly overstated in previous retail gravity research.

However, existing studies of retail gravity assume independence among errors after controlling for distance among customers and retailers. Since spatial dependence pervades other forms of real estate data, it seems reasonable to examine retail data for symptoms of

[1] Hardin and Wolverton (2001) provide evidence that retail gravitation affects rental rates.
[2] Examples of these studies are Gautschi (1981), Okoruwa, Terza, and Nourse (1988), and Okoruwa, Nourse, and Terza (1994), and Eppli and Shilling (1996).

spatial dependence. To do so, we estimate a retail gravity model with explicit spatial dependence.[3] In particular, we model the spatial dependencies with a spatial simultaneous autoregressive error process (Ord, 1975) among both consumers and retailers.

Our results confirm the importance of modeling both forms of spatial dependence in a retail-gravity model. When the spatial dependence is explicitly taken into account, the estimated parameters of variables pertaining to consumers and retailers change their magnitudes considerably, and many reverse their signs. In the case of store size, the parameter estimate goes from an implausible statistically significant negative value under OLS to a more plausible statistically significant positive value after allowing for spatial dependence. Compared with the spatial error model, OLS significantly underestimates the magnitude of the distance parameter for these data. Contrary to the suggestions by Gautschi (1981) as well as Eppli and Shilling (1996), the results show the importance of the distance parameter may be materially understated due to the inappropriate assumption of error independence. The inappropriate assumption of independence may lead to overestimation of the deterministic extent of trade areas for retail stores, and thus understate the importance of good locations.

The rest of this paper is organized as follows: Section 2 discusses retail gravitation and spatial dependence, section 3 describes retail sales data and empirical methods, section 4 presents empirical results; and section 5 concludes with the key results.

II. Retail Gravitation and Spatial Dependencies

Social scientists have drawn an analogy between the spatial interactions of individuals and Newton's law of gravity in physics. Over seventy years ago, Reilly (1931) formally applied the Newton's gravity concept to retail geography, and many models of shopping

[3] Porojan (2001) applied spatial statistics to estimate gravity models of international trade flows. Note, this provides a simpler scenario since each country is both an origin and destination. Thus, modeling spatial dependence for origins also models spatial dependence for destinations in trade studies.

behavior have been developed based on the concept of retail gravitation.[4] Many of these models relate the interaction (shopping trips or expenditures) between retail store b and consumer c (denoted by Z_{bc}) to the characteristics of store b (denoted by m_b), the characteristics of consumer c (denoted by m_c), and the separation measurement between b and c (denoted by d_{bc}) as in Equation (1),

$$Z_{bc} = \kappa m_b^{\beta_b} m_c^{\beta_c} d_{bc}^{\beta_d} \tag{1}$$

where κ is a constant, and β_b, β_c, and β_d are parameters to be estimated.[5]

Many earlier studies included only store size and distance in their gravity models.[6] Examples include Huff (1965) as well as Lakshmanan and Hansen (1965). Using survey data of shopping trips, Gautschi (1981) calibrated Huff's variation of Equation (1). He suggested that previous studies omitting other retail center variables might overstate the distance parameter in retail gravity models. Stanley and Sewall (1976) calibrated Huff's variation on single stores in a retail chain. Similar to Kolter (1971), Stanley and Sewall (1976) did not find that store size contributed significantly to estimates of store patronage. Both Stanley and Sewall (1976) as well as Kolter (1971) conclude that Huff's model is of limited value in estimating sales potential for single stores. Obtaining actual sales data, Eppli and Shilling (1996) calibrate Lakshmanan and Hansen's (1965) variant with an interactive approach and OLS (ordinary least squares). They find that store location (proximity to the competition) is of little importance and conclude that the distance parameter for retail gravity models may be significantly overstated.[7]

More recent research incorporates more characteristics of stores and consumers in retail gravity models. For example, Okoruwa, Terza, and Nourse (1988) include retail center variables such as age and type, as well as economic and demographic characteristics of

---

[4] See Brown (1992) for a list of studies of retail gravity models.

[5] Gravity models in this form can be applied to all sorts of spatial interaction behavior such as retail shopping, and population migration (Fotheringham and Webber, 1980).

[6] See Okoruwa, Terza, and Nourse (1988) for a list of such studies.

[7] In most retail gravity models, location means the distance among consumers and retailers.

shoppers in estimating shopping trip frequencies obtained from a survey. Okoruwa, Terza, and Nourse (1988) calibrate Equation (1) with a Poisson regression. Contradicting the typical hypothesis of retail gravity models, they find that retail center size exerts a negative influence on patronization rates.

While the gravity model incorporates space via a distance variable, this may not provide a sufficiently rich means of modeling interdependence among customers and retailers. For example, spatial dependence of errors among stores could arise due to omission of variables such as accessibility (e.g., turn lanes into centers or lights), visibility of signage, and retail demand externalities within a shopping center arising through clustering of stores. Clustering among heterogeneous retailers facilitates multi-purpose shopping behavior of consumers to reduce total travel costs, and clustering of homogeneous retailers facilitates comparison-shopping behavior (Eppli and Benjamin, 1994).[8] Studies have established the importance of these retailer and consumer behaviors in the choice of retail shopping trips (Eppli and Benjamin, 1994). Nevertheless, most of previous empirical studies did not incorporate these behaviors in retail gravity models.[9]

Spatial dependence of errors among consumers could arise due to clustering of consumers with similar circumstances (*e.g.*, individuals living in heavily shaded areas might not spend as much at garden centers), common traffic routes, and sharing of information.

Previous studies did not incorporate such dependence in retail gravity models. Modeling the dependence offers the possibility of better parameter estimation, correct inference (OLS standard errors are biased downwards in the presence of positive spatial dependence), and improved prediction (Cressie, 1993).

III. Retail Sales Empirical Data, Methods, and Model

[8] Most of gravity models assume that consumers shop from fixed points (*e.g.*, their places of residence) and buy just one type of good or service per shopping trip (Carter, 1993).
[9] One exception is Nevin and Houston (1980) who include multipurpose shopping opportunities in their gravity model.

Section A describes the retail sales data and census data employed in this study. Section B presents the SAR (simultaneous autoregressions) error model, section C provides information on the spatial dependence specification, section D gives the relevant likelihood function, while E shows the empirical model.

A. Retail Sales Data and Census Data

A retail consultant provided individual store and consumer data of a retail chain in the Houston market under a non-disclosure agreement preventing the release of specific data. The consumer data are for each household who shopped at a particular store. The variables are the total dollar amounts each household spent at each individual store for the year 2000, and the block group where each household resides. We aggregate the data to the block-group level and calculate retail sales in a block group for each store ($Sales_{bc}$). The individual store variables are total store sales in year 2000 (*store sales*) and in year 1999 (*lagged store sales*), store size in square feet (*store size*), type of shopping center where each store resides (*strip*, *pad*, or *mall*), age of the shopping center (*center age*), as well as longitude and latitude of each store.[10]

We supplement the retail sales data with 1990 census data and 1998 census estimates. In particular, we obtain census data relevant to the total potential expenditure for a block-group. The data are median medical supplies expenditure (*medical supp*), median household income (*med inc*), median house value (*med val*), median house age (*house age*), total population (*tot pop*), land area (*area land*), median age (*med age*), white population (*pop white*), and female population (*females*). Median house value, median house age, and land area are for 1990, while the other data are for 1998. We also obtain the longitude and latitude of each block-group to calculate the distances among retailers and consumers. Specifically, we compute the distance that consumers travel from their block group to the

[10] A strip (linear) shopping center consists of a line of stores with a pedestrian walk along the storefronts. A pad (cluster) center is a group of freestanding retail sites linked together by pedestrian walkways.

stores ($distance_{bc}$). Research has revealed that shopping trips may not be home based (Brown, 1992). Therefore, we also obtain average travel time to work ($travel\ time$) for 1990 from 1990 census data to take into account shopping trips originating from places of employment.

Geographically, the Houston market covers Baytown, Friendswood, Houston, Humble, Lake Jackson, Sugar Land, and Texas City. The retail chain has 14 stores in the Houston market. Twelve stores are located in shopping malls, one in a strip shopping center, and one in a pad-type shopping center.

Table 1 presents the descriptive statistics of the data used in this study. On average a store had $1,846 of retail sales in a block group. Retail sales in a block group vary widely from only $5 to $113,323. This variation indicates that retail sales do not uniformly originate over space. The distance among retailers and consumers is measured using the Euclidean metric. Over fifty percent of the consumers traveled less than 10.5 miles from their residence to the store where they shopped. Some consumers lived only 0.15 mile away from the store they patronized while others resided in distant locations.

For the 14 stores in our study, total store sales were stable over year 1999 and year 2000. A store on average generated about $1.3 million in annual sales. The stores generated sales between about $0.5 million and $2 million. The average store size was 4,579 square feet. The smallest store had 3,157 square feet, and the largest store had 6,612 square feet. The average shopping center age was 10 years. The newest center opened 5 years ago, and the oldest center opened 15 years ago.

There are 2,977 block-groups in Texas whose residents shopped at the stores in the study during 2000. Residents in a block group spent on average about $62,000 on medical supplies in 1998. The median household income was $45,707 a year. The median house value was about $71,000. Residents on average spent 25 minutes traveling from their home to work. On average a block group has 1,837 residents and 10 square kilometers in area. The median age of a resident was of 34 years old. A block group on average has 1,408 white residents and 928 female residents.

B. A SAR in Errors model

We describe a SAR in errors model, following the notation of Pace and Gilley (1997),

$$(I - \alpha D)Y = (I - \alpha D)X\beta + \varepsilon \tag{2}$$

where Y represents a $n \times 1$ vector of observations on the dependent variable, the matrix X contains *n* observations on *k* independent variables, α is a scalar parameter, D is an $n \times n$ spatial weight matrix, and ε is an $n \times 1$ vector of error terms. When errors exhibit spatial autocorrelation, a SAR in errors model partially differences each variable with the value of that variable at nearby observations (DY, DX). After transforming both sides, the errors are normal iid. This model is equivalent to: $Y = X\beta + \varepsilon$ where $\varepsilon \sim N(0, \Omega)$ and $\Omega^{-1} = (I - \alpha D)'(I - \alpha D)$.

In the SAR in error model, $\alpha > 0$ indicates a positive spatial dependence. This implies that errors of same sign are geographically clustered together. On the other hand, a $\alpha < 0$ implies a negative spatial dependence, and this implies that the errors of the opposite sign are clustered together geographically. When $\alpha = 0$, the SAR in errors model reduces to an OLS model. In a retail context, similar consumer populations and shopping environments may result in a positive spatial dependence among consumers. Individual stores sharing a similar retailing environment with other stores may lead to a positive spatial dependence among stores in a retail chain.

To prevent an observation from directly predicting itself, D has zeros on its diagonal. To facilitate interpretation, each row of D sums to 1. To ensure the stability of the entire error process, the spatial autocorrelation parameter, α, is restricted to be less than one, and since negative dependence seems unlikely, we restrict α to the interval [0,1).[11] These assumptions are summarized in the following:

[11] This assumption is for convenience. If $\alpha < 0$, the estimates should lie on the boundary $\alpha = 0$. We did not observe such a boundary solution.

$$
\begin{aligned}
&(a) \quad \underset{(n\times n)}{D}\underset{(n\times n)}{[1]} = \underset{(n\times n)}{[1]} \\
&(b) \quad diag(D) = \underset{(n\times n)}{[0]} \\
&(c) \quad 0 \le \alpha < 1 \\
&(d) \quad \varepsilon \sim N(0, \sigma^2 I)
\end{aligned} \tag{3}
$$

C. Specification of the Spatial Weight Matrix

To model the spatial dependence among stores and consumers, we specify a spatial weight matrix $D = wC + (1-w)S,$ where C and S are weight matrices for consumers and stores respectively, and $0 \le w \le 1.$ When $w = 1,$ D reduces to $C.$ When $w = 0,$ D reduces to $S.$ Empirically, we search for the optimal w over [0.00, 0.01,…1.00].

Since D is a $n \times n$ matrix, a straightforward dense specification of this would quickly result in intractable computations. Accordingly, we take a route that preserves sparsity among the components of D. This permits use of this approach in both large and small markets.

We form C using the approach of nearest neighbors with geometrically decaying weights. Under this scheme, the weight given to each customer depends on the proximity of each customer relative to all other customers. To make this feasible, we consider only the *m* nearest customers (nearest neighbors) to each customer. To make this more flexible, we allow for a geometrically declining weight for more distant neighbors.

Let $N^{(h)}$ represent an *n* by *n* matrix where $N_{ij}^{(h)} > 0$ when observation *j* is the *h*th nearest neighbors for observation *i* ($h = 1,2,\cdots m$, $i, j = 1,2,\cdots n$, $i \ne j$) and let $N_{ij}^{(h)} = 0$ otherwise. Thus, $N^{(1)}, N^{(2)}, \cdots, N^{(h)}$ represents a sequence of neighbor matrices. Let ρ represent the rate of geometric decay of weights such that the hth closest neighbors have a weight of ρ^h where $0 \le \rho \le 1.$ Define C as,

$$
C = \sum_{h=1}^{m} \rho^h N^{(h)} \Big/ \sum_{h=1}^{m} \rho^h
$$

and by construction each row sums to 1 and has zeros on the diagonal. Later, we search for the optimal C by varying m for 36 values over [1, 2…36] and varying ρ for 101 values over [0.00, 0.01…1.00].

Turning to means of modelling spatial dependence among stores, we create S as a product of sparse matrixes to overcome the computer memory requirement. We can divide this into five steps. The first step forms the Delaunay triangularization among stores. The Delaunay triangulation is the geometric dual of the Voronoi diagram that depicts the geometric expression of connections among contiguous stores (Calciu and Salerno, 1997).[12] Each store at a Delaunay triangle serves similar consumer populations who tend to live together. With this approach, let $V_{ij} = 1$ when observations i and j share contiguous triangles $(i \neq j)$ and let $V_{ij} = 0$ otherwise. Thus, V represents a $n_s \times n_s$ weight matrix of Delaunay triangulation for n_s stores.

The second step involves the row-standardization of V. Let

$$G_i = V_{ij} \Bigg/ \sum_{\substack{j=1 \\ i \neq j}}^{n_s} V_{ij}$$

and hence $G[1_{n_s}] = [1_{n_s}]$, where $[1_{n_s}]$ represents a vector of ones.

The third step involves the aggregation of customers at each store. Let $A_{ij} = 1$ when customer i shops at store j, and zero otherwise. Therefore, A is a 0, 1 matrix whose dimension is $n \times n_s$.

For the fourth step, we can standardize A' as well. Let R represent a $n_s \times n_s$ diagonal matrix with elements equal to the reciprocal of the sum of the columns of A. In which case, RA' will be row-stochastic, and thus $RA'[1] = [1]$. Now forming S as a product of these matrices,

$$S = AG(RA')$$

[12] The Voronoi diagram has the property that for each store every point in the region around that store is closer to that site than to any of the other stores.

yields a feasible means of quantifying spatial averages at the customer level.

To make this more intuitive, consider some variable $n \times 1$ vector v which might represent customer house values, a measure of wealth. The $n_s \times 1$ vector $RA'v$ computes the average customer house price for each store (aggregates from n customers to n_s stores), the $n_s \times 1$ vector $GRA'v$ computes the spatial average of customer house values at competing stores for each store, and the $n \times n_s$ matrix A redistributes the store level data back to the customer level. Thus, $AGRA'v$ represents the average house values of customers that shopped at nearby stores. If v denotes an error, $AGRA'v$ would measure the average errors at nearby stores for each customer. If the independent variables underpredicted the performance of nearby stores for some set of stores, $AGRA'v$ would be positive for customers who shopped at the store, and the use of this information could improve model predictions.

With this relationship, we can perform operations with S without needing to store an $n \times n$ matrix. Instead, for some v we can first form $RA'v$, a $n_s \times 1$ vector, multiply this by an $n_s \times n_s$ matrix, and then redistribute this result to each observation via the $n \times n_s$ matrix A. None of these operations requires much time or memory. As a result, in our actual computation, $D = wC + (1-w)AGRA'$ instead of $D = wC + (1-w)S$.

Here is a numerical example showing how to express S with G and A. Assume we have following matrixes for 4 retail sale observations for 3 stores:

$$A = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \; RA' = \begin{bmatrix} 0.5 & 0.5 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}, \; G = \begin{bmatrix} 0 & 1/2 & 1/2 \\ 1/2 & 0 & 1/2 \\ 1/2 & 1/2 & 0 \end{bmatrix},$$

then,

$$S=\begin{bmatrix} 0 & 0 & 0.5 & 0.5 \\ 0 & 0 & 0.5 & 0.5 \\ 0.25 & 0.25 & 0 & 0.5 \\ 0.25 & 0.25 & 0.5 & 0 \end{bmatrix}.$$

D. Maximum Likelihood Computations

The SAR in error model in (2) and (3) has the following profile log-likelihood function,

$$L(\alpha)=\kappa+\ln|I-\alpha D|-\frac{1}{2}\ln\left(SSE(\alpha)\right). \qquad (4)$$

where $SSE(\alpha)=(Y-X\beta)'(I-\alpha D)'(I-\alpha D)(Y-X\beta)$ and κ represents a constant (Pace, Barry, Slawson, and Sirmans, 2002). We maximize the log-likelihood by computing Equation (4) for 100 values of α over $[0.00,0.01,....0.99]$.

To overcome the computer memory requirement, we implement the log-determinant estimator of Barry and Pace (1999) to compute estimates of $\ln|I-\alpha D|$. In particular, $\ln|I-\alpha D|$ can be expanded in a power series as follows:

$$\ln|I-\alpha D|=\sum_{r=1}^{\infty}\frac{-\alpha^r\operatorname{tr}\left(D^r\right)}{r}$$

Using a finite approximation of the above expansion together with $E\left(\frac{u'Mu}{u'u}\right)=tr(M)$ for any real $n\times n$ matrix M and n by 1 column vector $u\sim N_n(0,I)$, Barry and Pace (1999) show a computationally simple way of estimating the log-determinant, and provide confidence limits for the estimated log-determinant.

Note, computation of the log-determinant only requires repeated multiplication of a D by a vector u since D^2u equals $D(Du)$, D^3u equals $D\left(D^2u\right)$, and so forth. For sparse D with number of elements proportional to n, each matrix-vector calculation uses $O(n)$ operations. Since $D=wC+(1-w)AGRA'$, Du just requires a series of low-cost computations, and this applies to $D^r u$ for $r=1\ldots q$ as well.

In terms of computing the Monte Carlo estimate of $\ln|I-\alpha D|$, we set $q=98$ and use 30 independent realizations of u.

E. The Empirical Model

We rewrite the retail gravity models, Equation (1), in log form and empirically model retail sales as:

$$\ln(Sales_{bc}) = X\beta + \varepsilon_{bc} \tag{5}$$

where $X = [X_{bc} \vdots X_b \vdots X_c]$, $X_{bc} = [1, \ \ln(distance_{bc})]$, X_b = a vector of variables pertaining to store b, X_c = a vector of variables pertaining to shoppers' area c, and ε_{bc} represents an error term.

The matrix X_b contains five variables of store characteristics. The five variables are $\ln(lagged\ store\ sales)$, $\ln(center\ age)$, $\ln(store\ size)$ as well as two store-type dummies, $strip$ and pad. The matrix X_c contains nine variables in log form relevant to total potential expenditure and average travel time to work in log form for block group c. Specifically, the 10 variables are $\ln(medical\ supp)$, $\ln(med\ inc)$, $\ln(med\ val)$, $\ln(house\ age)$, $\ln(tot\ pop)$, $\ln(area\ land)$, $\ln(med\ age)$, $\ln(pop\ white)$, $\ln(females)$, and $\ln(travel\ time)$.

To model the spatial dependence among stores and consumers, we fit Equation (5) with a SAR in error model using Pace and Barry's (2002) spatial statistics toolbox 1.1. Specifically we assume $\varepsilon_{bc} \sim N(0,[(1-\alpha D)'(1-\alpha D)]^{-1})$.

In summary, we hypothesize the signs for the variables in Equation (5) are positive signs for $\ln(lagged\ store\ sales)$, $\ln(store\ size)$, $\ln(medical\ supp)$, $\ln(med\ inc)$, $\ln(med\ val)$, $\ln(tot\ pop)$, $\ln(area\ land)$, $\ln(med\ age)$, $\ln(pop\ white)$, $\ln(females)$, and negative for $\ln(distance_{bc})$, $\ln(center\ age)$, $\ln(house\ age)$, $\ln(travel\ time)$, $strip$, and pad.

IV. Empirical Results

To understand the importance of spatial dependence, we calibrate four models that consider different components of spatial dependence among consumers and stores. The first model ignores spatial dependence. The second model considers spatial dependence among stores. The third model considers spatial dependence among consumers. The fourth model considers both spatial dependence among stores and consumers. Table 2 presents the calibration results of Equation (5).

The first model calibrates the gravity model with OLS. As hypothesized, the distance variable has a significant coefficient of –0.818, with signed root deviance (SRD) –57.346.[13] This coefficient is the constant distance elasticity of retail sales, which measures the proportional change in retail sales with respect to a small proportional change in distance. In particular, the coefficient here predicts a 0.818% decrease in a store's sales in a block group when the distance between the store and block groups increases 1%.

The strip shopping center indicator variable has a significant and negative coefficient, as hypothesized. This result matches those of Sirmans and Guidry (1993) as well as Oppewal and Timmermans (1999). A mall has more aesthetically appealing design and usually provides more protection to shoppers from the weather than other types of shopping centers (Sirmans and Guidry, 1993). Oppewal and Timmermans (1999) find that design influences consumer perception of shopping centers, and thus affects retail sales. However, several other variables have significant coefficients with signs opposite to those hypothesized by retail gravity. For variables pertaining to consumers, median household income and total population of a block group have significant and negative coefficients. Their signs are inconsistent with our hypotheses. For variables pertaining to retail stores, store size and shopping center age have significant coefficients with signs inconsistent with our hypotheses. Shopping center age has a positive coefficient while hypothesized to have a negative coefficient. In addition, store size has a negative coefficient while hypothesized to have a positive coefficient.

[13] The signed root deviance or SRD equals the square root of likelihood ratio statistic with a sign of its coefficient. It has an interpretation similar to a *t*-ratio. Squaring the SRD yields back the likelihood ratio statistic. We employ it to avoid scaling problems.

The second model only examines the spatial dependence among stores. The distance variable increases slightly from –0.818 to –0.815 and remains significant with a SRD –57.204. Although lagged store sales changes its coefficient from a negative value to the hypothesized positive value, the coefficient is insignificant. The dummy for a pad shopping center changes to have the hypothesized negative coefficient but remains insignificant. Store size still has a significant and negative coefficient. However, its SRD decreases near 43% in absolute value from –9.378 to –5.355. The spatial dependence among stores does not change either the signs or the significances of the coefficients of variables pertaining to consumers. Nevertheless, the signed root deviances for median household income and total population of a block group decrease 3.46% and 34.84% in absolute value respectively.

The third model considers only the spatial dependence among consumers. The distance estimated elasticity dramatically decreases by 68.34% (from –0.815 in the second model to –1.372 in the third model). The associated SRD increases more than 19% in absolute value from –57.204 to –68.304. In addition, spatial dependence among customers materially affects the estimated parameters. The coefficients of median household income and total population of a block group become insignificant, but still have signs opposite to our hypotheses. As a result, now there are no consumer variables having significant coefficients with signs opposite to our hypotheses. Median house age, median age of consumers, female population, and average travel time to work change to have insignificant coefficients. Nevertheless, median medical supplies expenditure changes to have a positive and significant coefficient as hypothesized.

The spatial dependence among consumers also influences store estimated parameters. Lagged store sales, a proxy for store management, changes its coefficient from insignificant to significant positive, as hypothesized. This coefficient is consistent with Black's (1966) argument. More sales enable a store to carry a greater variety of products, improve its services to customers, and compete with other stores. In addition, lagged store sales may capture the effect of store age on better management or other unobservables. Start-up problems may adversely affect sales for stores in new locations (Hise, Kelly, Gable, and McDonald, 1983). Stores in business longer should have overcome the start-up problems and have larger store sales. All these effects can increase the attractive power of a store. Store size still has a significant and negative coefficient. Nonetheless, it increases from –0.888 to –0.557, and its SRD decreases 1.21% in absolute value from –5.355 to –5.290. However, the dummy for a pad shopping center changes to have a significant and positive coefficient. This coefficient is not consistent with Sirmans and Guidry (1993) as well as Oppewal and Timmermans (1999).

The fourth model models both spatial dependence among stores and consumers. Compared to the third model, the distance elasticity slightly decreases another 0.36% from –1.372 to –1.377. The associated SRD increases another 2.07% in absolute value from –68.304 to –69.721. The next most important variable, store size, in retail gravity models now has a significant coefficient with hypothesized sign. Its coefficient changes to significant and positive in Model 4 from significant and negative in Models 1, 2, and 3. The coefficient changes from –0.557 in Model 3 to 0.490 in Model 4. The SRD for store size changes from –5.290 to 4.021. The dummy for a pad shopping center changes to have a significant and negative coefficient, as hypothesized. In addition, center age changes to have a negative coefficient. Although the coefficient is not significant, it has the hypothesized sign. This sign agrees with Sirmans and Guidry's (1993) as well as Gatzlaff, Sirmans, and Diskin's (1994) arguments. They argue that older shopping centers generally suffer functional or physical deficiencies and have an inappropriate tenant mix due to changing markets, and thus have less attractive power. The coefficient for lagged store sales increases about 58% in magnitude from 0.190 to 0.300. Its SRD also increases by over 50% from 4.644 to 7.012. The coefficients of total population of a block group become positive, as hypothesized, but remain insignificant. Median household income still has an insignificant and negative coefficient. However, this may not seem too surprising given the model already contains medical supply expenditures, a more direct measure of medical spending potential than income. Nevertheless, the coefficient decreases more than 97% in absolute value from –0.038 to -0.001. The SRD for the median household income also decreases more than 97% in absolute value from –0.895 to –0.024.

In addition to producing plausible estimated parameteers, model selection criteria also show models perform better when incorporating both forms of spatial dependence. The R^2 increases from 0.386 for the first model to 0.548 for the fourth model. The log-likelihoods also show that the fourth model outperforms the other three models. In fact, the likelihood ratio statistic between Model 1 and Model 4 is about 2,005.

V. Conclusions

Gravity-type models have been applied to the retail context extensively. However, the results have often seemed disappointing. For example, Kolter (1971) found that store size did not prove significant in predicting store sales. Both Stanley and Sewall (1976) as well as Kolter (1971) conclude the gravity model was of limited value in predicting single store sales. However, previous studies have assumed independence among customers and stores. For these data, modeling spatial dependence results in far more plausible parameter estimates than assuming independence. For example, under independence previous store sales have an insignificant, negative effect on future store sales. After modeling spatial dependence, previous store sales have significant, positive effect on future store sales. Under independence, store size has a significant, negative effect on store sales. After modeling spatial dependence, store size has a significant, positive effect on store sales. Under independence, center age has a significant, positive effect on store sales. After modeling spatial dependence, center age has a insignificant, negative effect on store sales.

Even the distance variable, the central feature of the gravity model, may not be well estimated under independence. Using actual sales for a retail chain in the Houston market and modeling spatial dependence among customers and stores, we find that the assumption of independent errors can lead to understating the magnitude of the distance parameter by as much as two-thirds. This result implies that previous studies may have overestimated the deterministic extent of trade areas as reflected in the distance parameter, and thus have understated the importance of good locations.

References

Barry, Ronald Paul and R. Kelley Pace. (1999). "Monte Carlo Estimates of the Log Determinant of Large Sparse Matrices," *Linear Algebra and Applications* 289, 41-54.

Black, J. (1966). "Some Retail Sales Models," Paper for the Urban Studies Conference, Oxford.

Brown, S. (1992). "The Wheel of Retail Gravitation?" *Environment and Planning A* 24, 1409-1429.

Calciu, M. and F. Salerno. (1997). "A New Approach to Spatial Management of Retail Networks, Based on German School's Central Place Theory: Application to Bank Location," dans H. Mühlbacher /J-P Flipo (Eds.), *Advances in Services Marketing.* Gabler, Wiesbaden.

Carter, Charles C. (1993). "Assumptions Underlying the Retail Gravity Model," *The Appraisal Journal* 4, 509-518.

Cressie, Noel. (1993), *Statistics for Spatial Data*, New York: Wiley.

Eppli, Mark J. and John D. Benjamin. (1994). "The Evolution of Shopping Center Research: A Review and Analysis," *Journal of Real Estate Research* 9, 5-32.

Eppli, Mark J. and James D. Shilling. (1996). "How Critical Is a Good Location to a Regional Shopping Center?" *Journal of Real Estate Research* 12, 459-468.

Fotheringham, A. Stewart and M. J. Webber. (1980). "Spatial Structure and the Parameters of Spatial Interaction Models," *Geographic Analysis* 12, 33-46.

Gatzlaff, D. H., G. S. Sirmans, and B. A. Diskin. (1994). "The Effect of Anchor Tenant Loss on Shopping Center Rents," *Journal of Real Estate Research* 9, 99-110.

Gautschi, David A. (1981). "Specification of Patronage Models for Retail Center Choice," *Journal of Marketing Research* 18, 162-174.

Hardin, William G. III and Marvin L. Wolverton. (2001). "Neighborhood Center Image and Rents," *Journal of Real Estate Finance and Economics* 23, 31-46.

Hise, Richard T., J. Patrick Kelly, Myron Gable, and James B. McDonald. (1983). "Factors Affecting the Performance of Individual Chain Store Units: an Empirical Analysis," *Journal of Retailing* 59, 22-39.

Huff, D. L. (1965). "A Probabilistic Analysis of Shopping Center Trade Area," *Land Economics* 39, 81-90.

Kolter, Philip. (1971). *Marketing Decision Making: A Model Building Approach*, Holt, Rinehart & Winston, New York.

Lakshmanan, T. R. and Walter G. Hansen. (1965). "A Retail Market Potential Model," *Journal of the American Institute of Planner* 13, 134-43.

Mejia, Luis C. and John D. Benjamin. (2002). "What Do We Know About the Determinants of Shopping Center Sales? Spatial vs. Non-Spatial Factors," *Journal of Real Estate Literature* 10, 3-26.

Nevin, J. R. and M. J. Houston. (1980). "Image as a Component of Attraction to Intraurban Shopping Areas," *Journal of Retailing* 56, 77-93.

Okoruwa, A. Ason, Joseph V. Terza, and Hugh O. Nourse. (1988). "Estimating Patronization Shares for Urban Retail Centers: An Extension of the Poisson Gravity Model," *Journal of Urban Economics* 24, 241-259.

Okoruwa, A. Ason, Hugh O. Nourse, and Joseph V. Terza. (1994), "Estimating Sales for Retail Centers: An Application of the Poisson Gravity Model," *Journal of Real Estate Research* 9, 85-97.

Oppewal, H. and H. Timmermans. (1999). "Modeling Consumer Perception of Public Space in Shopping Centers," *Environment and Behavior* 31, 45-65.

Ord, J. K. (1975). "Estimation Methods for Models of Spatial Interaction," *Journal of the American Statistical Association* 70, 120-126.

Pace, R. Kelley, Ronald Barry, V. Carlos Slawson, Jr., and C. F. Sirmans. (2002). "Simultaneous Spatial and Functional Form Transformations," in Luc Anselin and Raymond Florax (Ed.), *Advances in Spatial Econometrics,* forthcoming, Springer-Verlag, Heidelberg.

Pace, R. Kelley and Otis W. Gilley. (1997). "Using the Spatial Configuration of the Data to Improve Estimation," *Journal of Real Estate Finance and Economics* 14, 333-340.

Pace, R. Kelley and Ronald Barry. (2002). Spatial Statistics Toolbox 1.1. Available: www.spatial-statistics.com.

Porojan, A. (2001). "Trade Flows and Spatial Effects: The Gravity Model Revisited," *Open Economic Review* 12, 265-280.

Reilly, W. J. (1931). *The Law of Retail Gravitation*, William J. Reilly Co., New York.

Sirmans, C. F. and Krisandra A. Guidry. (1993). "The Determinants of Shopping Center Rents," *Journal of Real Estate Research* 8, 107-115.

Stanley, Thomas J. and Murphy A. Sewall. (1976). "Image Inputs to a Probabilistic Model: Predicting Retail Potential," *Journal of Marketing* 40, 48-53.

Table 1: Descriptive Statistics

Variables	Label	Mean	Std Dev.	Median	Minimum	Maximum
Retail sales ($)	$sales_{bc}$	1,846.06	3,827.73	524.00	5.00	113,323.00
Distance (miles)	$distance_{bc}$	22.29	45.99	10.49	0.15	712.26
Store sales ($)	*store sales*	1,339,666.71	380,706.15	1,315,324.50	505,601.00	2,030,448.00
Lagged store sales ($)	*lagged store sales*	1,324,174.57	383,531.07	1,268,676.50	515,538.00	1,853,128.00
Shopping center age (years)	*center age*	10.36	3.12	10.88	4.87	15.75
Store size (square feet)	*store size*	4,578.79	929.01	4,511.00	3,157.00	6,612.00
Median Medical supplies expenditure ($1,000)	*medical supp*	61.67	42.60	51.00	3.00	782.00
Median household income ($)	*med inc*	45,706.62	24,808.02	39,328.00	11,976.00	150,000.00
Median house value ($)	*med val*	71,301.38	57,253.59	57,100.00	14,999.00	500,001.00
Median house age (years)	*house age*	31.98	11.94	29.00	10.00	61.00
Travel time to work (minutes)	*travel time*	24.53	6.28	24.40	6.90	54.10
Total Population (persons)	*tot pop*	1,837.00	1,007.17	1,627.00	70.00	11,125.00
Land area (0.001 square kilometers)	*area land*	10,428.72	33,201.48	1,125.00	60.00	446,589.00
Median age (year)	*med age*	33.57	6.50	32.70	11.20	75.00
White population (persons)	*pop white*	1,407.99	876.56	1,282.00	5.00	10,135.00
Female population (persons)	*females*	927.87	507.12	826.00	43.00	6,019.00

Table 2: Retail Gravity Models

Independent variables	Model 1 β_{OLS}	Model 1 SRD	Model 2 β_{SAR}	Model 2 SRD
$\ln(distance_{bc})$	-0.818	-57.346^{+++}	-0.815	-57.204^{+++}
$\ln(lagged\ store\ sales)$	-0.007	-0.154	0.054	0.761
$\ln(center\ age)$	0.565	9.638$^{\times\times\times}$	0.514	7.529$^{\times\times\times}$
$\ln(store\ size)$	-1.059	-9.378$^{\times\times\times}$	-0.888	-5.355$^{\times\times\times}$
$strip$	-0.310	-8.947^{+++}	-0.305	-8.888^{+++}
pad	0.006	0.126	-0.016	-0.258
$\ln(medical\ supp)$	-0.030	-0.495	-0.029	-0.470
$\ln(med\ inc)$	-0.113	-3.123$^{\times\times\times}$	-0.109	-3.015$^{\times\times\times}$
$\ln(med\ val)$	0.144	3.581^{+++}	0.153	3.744^{+++}
$\ln(house\ age)$	-0.358	-7.201^{+++}	-0.346	-6.926^{+++}
$\ln(tot\ pop)$	-0.360	-2.586$^{\times\times\times}$	-0.252	-1.685$^{\times}$
$\ln(area\ land)$	0.189	17.861^{+++}	0.186	17.476^{+++}
$\ln(med\ age)$	0.236	2.425^{++}	0.229	2.349^{++}
$\ln(pop\ white)$	0.070	4.330^{+++}	0.064	3.902^{+++}
$\ln(females)$	0.509	3.910^{+++}	0.408	2.868^{+++}
$\ln(travel\ time)$	-0.683	-11.122^{+++}	-0.683	-11.117^{+++}
Intercept	12.700	8.802***	13.683	6.016***
w			0	
α			0.040	1.337
m			34	
ρ			1	
p			30	
q			98	
n	7,983		7,983	
k	17		20	
Log likelihood	-36,165.975		-36,165.081	
R^2	0.386		0.386	

1. w is the weight on the spatial weight matrix for consumers.
2. α is the spatial autoregression parameter.
3. m is the number of nearest blocks used in constructing the spatial weight matrix for consumers.
4. ρ represents 1 minus the rate of geometric decay of weights on neighboring blocks in the spatial weight matrix.
5. p is the number of trials performed to obtain the Monte Carlo estimates for the log-determinant term.
6. q is the highest order of power series expansion for the log-determinant term.
7. SRD stands for signed root deviance that equals the square root of likelihood statistics with a sign of its coefficient.
8. * significant at 10% level; ** significant at 5% level; *** significant at 1% level of sign without hypothesis.
9. $^{\times}$ significant at 10% level; $^{\times\times}$ significant at 5% level; $^{\times\times\times}$ significant at 1% level of sign counter to hypothesis.
10. $^{+}$ significant at 10% level; $^{++}$ significant at 5% level; $^{+++}$ significant at 1% level of sign consistent with hypothesis.

(**Table 2 continued)**

	Model 3		Model 4	
Independent variables	β_{SAR}	SRD	β_{SAR}	SRD
$\ln(distance_{bc})$	-1.372	-68.304^{+++}	-1.377	-69.721^{+++}
$\ln(lagged\ store\ sales)$	0.190	4.644^{+++}	0.300	7.012^{+++}
$\ln(center\ age)$	0.262	4.958$^{\times\times\times}$	-0.079	-1.394
$\ln(store\ size)$	-0.557	-5.290$^{\times\times\times}$	0.490	4.021^{+++}
$strip$	-0.251	-8.424^{+++}	-0.148	-5.065^{+++}
pad	0.096	2.347$^{\times\times}$	-0.112	-2.736^{+++}
$\ln(medical\ supp)$	0.431	6.049^{+++}	0.386	5.537^{+++}
$\ln(med\ inc)$	-0.038	-0.895	-0.001	-0.024
$\ln(med\ val)$	0.186	4.101^{+++}	0.198	4.398^{+++}
$\ln(house\ age)$	-0.019	-0.292	-0.051	-0.839
$\ln(tot\ pop)$	-0.170	-1.184	0.003	0.021
$\ln(area\ land)$	0.031	2.259^{++}	0.049	3.390^{+++}
$\ln(med\ age)$	0.010	0.098	0.034	0.352
$\ln(pop\ white)$	0.078	3.282^{+++}	0.064	2.755^{+++}
$\ln(females)$	0.163	1.228	0.044	0.333
$\ln(travel\ time)$	-0.104	-1.332	-0.077	-0.997
Intercept	7.782	5.296***	2.518	1.573
w	1		0.840	
α	0.870	42.409^{+++}	0.990	44.777^{+++}
m	34		34	
ρ	1		1	
p	30		30	
q	98		98	
n	7,983		7,983	
k	20		21	
Log likelihood	-35,264.586		-35,163.504	
R^2	0.540		0.548	

1. w is the weight on the spatial weight matrix for consumers.
2. α is the spatial autoregression parameter.
3. m is the number of nearest blocks used in constructing the spatial weight matrix for consumers.
4. ρ represents 1 minus the rate of geometric decay of weights on neighboring blocks in the spatial weight matrix.
5. p is the number of trials performed to obtain the Monte Carlo estimates for the log-determinant term.
6. q is the highest order of power series expansion for the log-determinant term.
7. SRD stands for signed root deviance that equals the square root of likelihood statistics with a sign of its coefficient.
8. * significant at 10% level; ** significant at 5% level; *** significant at 1% level of sign without hypothesis.
9. $^{\times}$ significant at 10% level; $^{\times\times}$ significant at 5% level; $^{\times\times\times}$ significant at 1% level of sign counter to hypothesis.
10. $^{+}$ significant at 10% level; $^{++}$ significant at 5% level; $^{+++}$ significant at 1% level of sign consistent with hypothesis.