Data-guided Gravity Model for Competitive Facility Location

Dawit Zerom
California State University-Fullerton
Zvi Drezner
zdrezner@fullerton.edu
California State University-Fullerton

Research Article

Keywords: Competitive facility location, Gravity model, Decay function, Data-guided

Posted Date: February 13th, 2024

DOI: https://doi.org/10.21203/rs.3.rs-3935429/v1

Additional Declarations: No competing interests reported.

Version of Record: A version of this preprint was published at Networks and Spatial Economics on March 27th, 2024. See the published version at https://doi.org/10.1007/s11067-024-09623-5.

Data-guided Gravity Model for Competitive Facility Location

Dawit Zerom and Zvi Drezner
College of Business and Economics,
California State University-Fullerton,
Fullerton, CA 92834.
e-mail: dzerom@fullerton.edu; zdrezner@fullerton.edu

Abstract

In this paper we introduce a data analytics approach for specifying the gravity model as applied to competitive facility location. The gravity model is used primarily by marketers to estimate the market share attracted by competing retail facilities. Once the market share is computed, various solution techniques can be applied for finding the best locations for one or more new facilities. In competitive facility location research, various parametrized gravity models have been proposed such as the power and the exponential distance decay specifications. However, parameterized approaches may not be robust to slight data inconsistency and possibly leading to inaccurate market share predictions. As the volume of data available to support managerial decision making is growing rapidly, non-parametric (data-guided) approaches are naturally attractive alternatives as they can mitigate parametric biases. We introduce a unified gravity model that encompasses practically all existing parametric gravity models as special cases. We provide a statistical framework for empirically estimating the proposed gravity models focusing on shopping malls data involving shopping frequency.

Key words: Competitive facility location, Gravity model, Decay function, Data-guided.

1 Introduction

Competitive facilities location problems attempt to find the locations for one or more new facilities among existing competing facilities that maximize the captured market share. The facilities attract demand generated by customers in the area. In most applications profit is increasing when market share captured increases. Therefore, the common objective is to maximize the market share captured by the new facilities. In the basic models such as locating shopping malls, grocery stores, furniture stores, restaurants, and many others, it is assumed that facilities cannot be easily moved once they were established. Such basic models were extended in many ways. For example, incorporating future changes in market conditions, have an available budget to be spent in building new facilities and/or improving the attractiveness of existing facilities by lowering prices, expanding offerings, improving appearance, or other investments. For a recent review of competitive models see Drezner and Eiselt (2023) which includes over 100 relevant references.

Competitive models require an accurate estimate of the captured market share at a given location. Then, optimization methods are applied to find the best location that maximizes the captured market share. There is no established rule how to reliably estimate the market share captured and many models were proposed for such estimation (Drezner and Eiselt, 2023). The most commonly investigated approach is the gravity model (for example, papers before the year 2000: Bell et al., 1998; Downs, 1970; Drezner, 1994; Jain and Mahajan, 1979; Nakanishi and Cooper, 1974; Prosperi and Schuler, 1976; Schuler, 1981; Timmermans, 1982, 1988), and more recently Aboolian et al. (2007a,b); Eiselt et al. (2015); Fernández et al. (2007). In gravity models, it is assumed that the proportion of the buying power attracted to a facility is proportional to the facility's attractiveness and to a distance decay function because the appeal of a facility decreases if the distance to it increases. This proportion can be interpreted as the probability that a customer selects a certain facility. Therefore, the sum of these proportions is equal to 1.

Suppose that n demand points exist in the area and demand point i has a buying power B_i. There are p facilities located in the area and facility j has an attractiveness level A_j. The distance between demand point i and facility j is d_{ij}. The market share M_j captured by facility j is

$$M_j = \sum_{i=1}^{n} B_i \frac{A_j f(d_{ij}, \lambda)}{\sum\limits_{k=1}^{p} A_k f(d_{ik}, \lambda)} \tag{1}$$

where $f(d, \lambda)$ is a distance decay function with a parameter λ. Note that $\sum\limits_{j=1}^{p} M_j = \sum\limits_{i=1}^{n} B_i$ which means that all the buying power is captured by the competing facilities. There are models that assume that some of the demand is lost and therefore the sum of the proportions is less than 1 (Berman et al., 2006; Drezner and Drezner, 2008, 2012).

Reilly (1931), who proposed the gravity model, suggested the distance decay function $\frac{1}{d^2}$ which imitates gravitational force, and coined the name "gravity model". For a detailed discussion of various distance decay functions see Drezner and Eiselt (2023).

The paper is organized as follows. In Section 2 a general formulation of an extended gravity model is introduced and analyzed. Various distance decay functions that do not depend on the attractiveness of the facilities, are proposed and analyzed in Section 3. Distance decay functions that depend on the attractiveness level are investigated in Section 4. In Section 5, the location of a new facility for different parameters is found and the various results discussed. We summarize and conclude the paper in Section 6.

2 A General Gravity Model

The original gravity model was recently extended in many ways (Drezner et al., 2020, 2022). We present the most recent extension (Drezner and Zerom, 2023b) that incorporates all the recent ones.

There are n demand points located in the area and p facilities. Some of them can be the new ones to be established. Define:

B_i	is the buying power at demand point i for $i = 1, \ldots, n$.
A_j	is the attractiveness level of facility j for $j = 1, \ldots, p$.
d_{ij}	is the distance between demand point i and facility j.
μ_{ij}	facility j's expected market share of the buying power at demand point i.
μ_j	the total expected market share captured by facility j.

Following the extended gravity model of Drezner and Zerom (2023b), μ_j is obtained by

$$\mu_j = \sum_{i=1}^{n} \mu_{ij} = \sum_{i=1}^{n} B_i \frac{f(d_{ij}, A_j)}{\sum\limits_{k=1}^{p} f(d_{ik}, A_k)} \tag{2}$$

where $f(d_{ij}, A_j)$ is defined by parameters $\alpha \geq 0$, $\beta \geq 0$, $\lambda \geq 0$ and a function $\phi(\cdot)$,

$$f(d, A_j) = e^{-\lambda\phi(d)+(\alpha+\beta\phi(d))\ln A_j}. \tag{3}$$

The above gravity model encompasses most models used in competitive facility location studies as special cases. For example, the restrictions $\alpha = 1$ and $\beta = 0$ are typically assumed. The function $\phi(d) = \ln d$ leads to the commonly used power distance decay function $\frac{1}{d^\lambda}$ originally proposed by Huff (1964, 1966). The exponential distance decay function $e^{=\lambda d}$ proposed by (Hodgson, 1981; Wilson, 1976) is obtained when $\phi(d) = d$. Bell et al. (1998) suggested $\phi(d) = d^\gamma$ for $\gamma \geq 0$. Using more than 30 thousand shopping trips to grocery stores, they found that $\gamma = 0.409$ fits their data the best.

The assumption $\beta = 0$ implies that all facilities share a common distance decay parameter $-\lambda$ regardless of their attractiveness level. If this assumption is relaxed, it is easy to see from (3) that the distance decay parameter is facility-dependent, i.e. is given by $-\lambda + \beta \ln A_j$. Therefore, a more attractive (appealing) facility will have a slower distance decay due to the moderation ($\beta > 0$) by its higher attractiveness A_j. Using shopping malls trips data and for both power and exponential distance decays, Drezner and Zerom (2023b) showed that the restriction $\beta = 0$ is not tenable and can lead to a significant loss in market share prediction accuracy.

Despite the empirical success of (3) relative to the more restrictive special cases, it is still a highly parameterized approach that may not be robust to slight data inconsistency. Such bias may, for example, lead to inaccurate market share predictions. To circumvent this we propose a data-driven (non-parametric) alternative to (3) where we make no assumptions on functional forms. Specifically, we introduce two extensions. The first is more general and is given by

$$f_j(d) = e^{g_j(d)} \tag{4}$$

where $g_j(\cdot)$ is an unknown smooth function representing facility-specific distance decay. Recently, Drezner et al. (2020) suggested a more restrictive special case of (4) where

$$g_j(d) = -\lambda_j \phi(d) \tag{5}$$

with each facility having its own distance decay parameter λ_j. The volume of data available to support managerial decision making is growing rapidly as firm operations become digitized. For example, retailers such as shopping malls or grocery stores routinely collect information from their customers such as their zip codes (that helps to determine the distance) and the frequency of their shopping trips. With just these two pieces of data, one can easily obtain a non-parametric estimate of equation (4) without imposing any unwarranted restrictions such as those in equation (5). Such data analytics can offer quick valuable actionable insight to management regarding the impact of location of facilities on customers' shopping frequency.

The second extension specifically assumes that any distance decay variation among facilities is explained by their respective attractiveness, i.e.

$$f(d, A_j) = e^{g(d, A_j)} = e^{g_0(d) + g_1(d) \ln A_j} \tag{6}$$

where both $g_0(\cdot)$ and $g_1(\cdot)$ are unknown smooth functions. Note that (6) is a direct extension of (3). The function $g_0(\cdot)$ can be viewed as the pure distance decay (common to all facilities) while $g_1(\cdot)$ as the decay moderator function. This way, we allow facility attractiveness A_j to moderate distance decay where the degree of moderation itself may vary with distance. For example, in (3) the degree of moderation is a constant given by β.

The proposed extensions (4) and (6) can be used by practitioners as a guide while searching for a more proper tractable parametric specifications or empirically validate currently used parametric models. In addition to its robustness, the extensions pose no estimation difficulty. While (3) can be estimated within the generalized linear models (GLM) framework (see Drezner and Zerom, 2023b), we introduce a localized GLM to estimate (4) and (6) that can be easily implemented in computational platforms such as R or Python.

3 Distance Decay Functions

Before the paper by Drezner et al. (2020), most competitive location papers assumed that the distance decay function is the same for all facilities, and concentrated on approaches to estimate the attractiveness levels of the facilities. Huff (1964, 1966) assumed that the attractiveness is proportional to the store area. Nakanishi and Cooper (1974) suggested that the attractiveness is proportional to a product of characteristics and each characteristic is rated by public opinion surveys. Drezner (2023) lists dozens of papers and attractiveness components that were investigated However, estimating the expected market share by actual customer behavior is more accurate than relying on public opinion surveys.

The purpose of this section is to empirically validate various parametric distance functions proposed in the literature assuming that the distance decay function does not depend on the facility attractiveness. Thus, from (6),

$$f(d_{ij}, A_j) = e^{g(d_{ij})} \tag{7}$$

where $g(\cdot)$ is unknown smooth distance decay function. If sufficient or dense data are available from most (or all) demand points or zip codes, one can also estimate a facility-dependent distance decay

$f_j(d) = e^{g_j(d)}$ by equation (7) by solely focusing on one facility at a time. In this case, we are no longer assuming facilities are equally appealing and in fact the resulting estimates are even more general than those by equation (6). Because the Orange County shopping malls data is very sparse (even though more than 3000 customers were intercepted in the seven malls) in the sense that there are several zero trips from demand point i to facility j, the facility-dependent distance decay $f_j(d)$ estimates are very likely to be highly unreliable. We would like to point out that the estimation procedure discussed in this section is versatile in the sense that it is applicable for estimating the common distance decay $f(d) = e^{g(d)}$ as in (7) when several facilities are jointly considered (the focus of this section) or the more general facility dependent decay $f_j(d)$ when estimation (with sufficient data availability) is done one facility at a time. For large retailers that attract customers from large geographic areas, their shopping frequency data will be sufficiently dense to allow reliable estimation of $f_j(d)$.

In the competitive facility location literature, it is traditionally assumed that $g(\cdot)$ belongs to a known parametric family of functions. In column 2 of Table 1, a variety of parametric models that will be explored in this paper are depicted. Note that one can also specify other plausible parametric functions within the framework of equation (7). Type 1 is similar to Bell et al. (1998), Type 2 is exponential decay, and Type 5 is power decay. Types 1- 6 are a single decay parameter models. To capture a more elaborate decay pattern, one may mix a pair of models from Types 1-6. Accordingly, Types 7 and 8 are constructed. Other pairs may also be combined.

Table 1: Pure Parametric Distance Decays.

Type	$g(d)$	$\hat{g}(d)$	$\hat{c}$	R^2
1	$\lambda_1 d^{0.5}$	$-1.13d^{0.5}$	-6.20	53%
2	$\lambda_2 d$	$-0.23d$	-7.43	50%
3	$\lambda_3 d^{1.5}$	$-0.05d^{1.5}$	-7.90	45%
4	$\lambda_4 d^2$	$-0.0125d^2$	-8.19	40%
5	$\lambda_5 \ln(d)$	$-1.16\ln(d)$	-7.05	50%
6	$\lambda_6 \ln^2(d)$	$-0.43\ln^2(d)$	-7.54	52%
7	$\lambda_7 d + \lambda_8 \ln(d)$	$-0.11d - 0.64\ln(d)$	-7.16	52%
8	$\lambda_9 d^{0.5} + \lambda_{10} \ln^2(d)$	$-0.84d^{0.5} - 0.11\ln^2(d)$	-6.54	53%

3.1 Estimation

We outline the estimation procedure of the distance decay function $f(\cdot)$ in (7) within the context of the gravity model in (2). First, we discuss how generalized linear models (GLM) can be used to estimate the parametric distance decays (Types 1 - 8) in Table 1. Later we extend GLM into a localized GLM to obtain $f(\cdot)$ via non-parametrically estimating $g(\cdot)$. To illustrate, we use mall shoppers survey data in Orange County, California (Drezner, 2006). As a surrogate for expected market share in (2), we consider

μ_{ij}: expected number of shoppers in mall j from zip code i.

A total of 80 zip codes ($1 \leq i \leq 80$) and 7 shopping malls ($1 \leq j \leq 7$) are considered. Using several aspects of the malls' appeal, Drezner (2006) also measured the attractiveness A_j of the seven malls. These attractiveness values will be incorporated into distance decay function estimation in Section 4. The buying power B_i at demand point i is represented by the population size at zip code i which is publicly available. The distance d_{ij} is based on the Euclidean distance where the latitude and longitude of the centers of zip codes were first converted to miles.

We assume that the number of shoppers from zip code i to mall j, denoted by y_{ij}, follows a Poisson probability distribution, i.e.

$$y_{ij} \sim Poisson(\mu_{ij}) \tag{8}$$

where

$$\mu_{ij} \equiv E(y_{ij}) > 0. \tag{9}$$

Note from (2) that μ_{ij}, is proportional to

$$\mu_{ij} \propto B_i \times e^{g(d_{ij})} \tag{10}$$

where $f(d_{ij}, A_j)$ uses (7). Now, given the observed data on B_i and d_{ij}, and after applying logarithms to both sides, we obtain the statistical model

$$\ln(\mu_{ij}) = c + \ln(B_i) + g(d_{ij}) \tag{11}$$

where c is an unknown constant of proportionality. Modeling the log-transformed $\ln(\mu_{ij})$ instead of μ_{ij} leads to linear parametric decay models (see Table 1) in their respective target parameters, i.e.

$$\ln(\mu_{ij}) = \ln(B_i) + c + \lambda X_{ij} \tag{12}$$

where for example $X_{ij} = d_{ij}^{0.5}$ for Type 1 decay and $X_{ij} = [\ln(d_{ij})]^2$ for Type 6 decay. For model Type 7,

$$\ln(\mu_{ij}) = \ln(B_i) + c + \lambda_1 X_{1,ij} + \lambda_2 X_{2,ij} \tag{13}$$

where $X_{1,ij} = d_{ij}$ and $X_{2,ij} = \ln(d_{ij})$. Parameter estimates (both the respective decay parameters and the constant c are given in Table 1). Specifically, we use the GLM routine (mgcv library) within the open source software R; See McCullagh and Nelder (2019) for a comprehensive review of GLM. In GLM, the target variable y_{ij} does not have to be normally distributed. GLM permits the distribution to be from an exponential family such as binomial, Poisson, Negative-Binomial, etc. and the unknown parameters c and λ are estimated using maximum likelihood. Note that the coefficient of $\ln(B_i)$ is forced to be 1. This implies that we are implicitly modeling the logarithm of "rate" of expected shoppers, i.e. $\ln(\mu_{ij}/B_i)$. This is an important feature as the number of shoppers originating from a zip code is expected to be proportional to the population size B_i.

3.1.1 Non-parametric $g(d)$

In this sub-section we discuss a non-parametric approach to estimate $g(\cdot)$ in (11) without assuming a parametric form (such as those in Table 1). Assuming that the function $g(d_{ij})$ is smooth, it follows from Taylor expansion (of order 1) that

$$g(d_{ij}) \approx g(d) + g'(d)(d_{ij} - d) = \theta_0 + \theta_1 Z_{ij} \tag{14}$$

where $Z_{ij} = d_{ij} - d$ and θ_0 includes the constant c. One may also consider polynomial orders higher than 1 to capture more complex non-linearity. Using (14), a localized version of (11) is given by

$$\ln(E(y_{ij}|d_{ij} = d)) = \ln(B_i) + \theta_0 + \theta_1 Z_{ij} \tag{15}$$

which is still linear, albeit locally, in the parameters θ_0 and θ_1. Using the Poisson distribution, the local likelihood function for $n \times p$ observed data points is

$$\prod_{i=1}^{n}\prod_{j=1}^{p}\frac{(B_i e^{\theta_0+\theta_1 Z_{ij}})^{y_{ij}}}{y_{ij}!}e^{-B_i e^{\theta_0+\theta_1 Z_{ij}}}. \tag{16}$$

Estimates for θ_0 and θ_1 are obtained by maximizing a "locally weighted" logarithm of (16) while ignoring terms that do not depend on the parameters, i.e.

$$[\hat{\theta}_0(d), \hat{\theta}_1(d)] = \underset{\theta_0,\theta_1}{\text{argmax}} \sum_{i=1}^{n}\sum_{j=1}^{p} K_h(Z_{ij})\Big(y_{ij}(\theta_0+\theta_1 Z_{ij}) - B_i e^{\theta_0+\theta_1 Z_{ij}}\Big). \tag{17}$$

The local weights $K_h(Z_{ij})$ satisfy $K_h(u) = h^{-1}K(u/h)$ for some $h > 0$ (bandwidth) where K is a kernel function that satisfies $\int K(t)dt = 1$; See for example Wand and Jones (1994) for a comprehensive review of kernel-based smoothing. In this paper, we use $K(t) = 0.75(1-t^2)\mathbf{1}_{|t|<1}$ and the bandwidth $h = 2\hat{\sigma}_d(n \times p)^{-1/5}$ where $\hat{\sigma}_d$ is the observed standard deviation of the distances. Note that in (17), log-likelihood value of those observations nearby (controlled by h) the distance point d is given more weight via $K_h(Z_{ij})$. Finally, the non-parametrically estimated decay function is given by

$$\hat{f}(d) = e^{\hat{\theta}_0(d)} = e^{\hat{g}(d)}. \tag{18}$$

We note that when estimating the market share captured by different facilities, the different attractiveness levels of the facilities are incorporated. In Table 1, $\hat{f}(d)$ is multiplied by A_j.

3.2 Performance of Parametric Decays

In Table 1, we report the estimated results for eight parametric decay models. In Figure 1, the estimated parametric decays are plotted against the non-parametric estimated decay (a proxy for the "true" decay). Types 1, 6, 7, and 8 appear to closely approximate the true decay. The R^2 indicates how well the parametric models predict the expected number of shoppers to malls. Type 1 and Type 8 are the best performers by their R^2 followed by Types 6 and 7. Overall, factoring the number of model decay parameters, one may conclude that Types 1 and 6 performed best overall. In competitive facility location studies, Type 2 and Type 5 decays are routinely used as the preferred distance decay patterns. Our analysis shows that this may not be necessarily the case and non-parametric (data-driven) approaches can be quite useful in guiding selection of more fitting decay

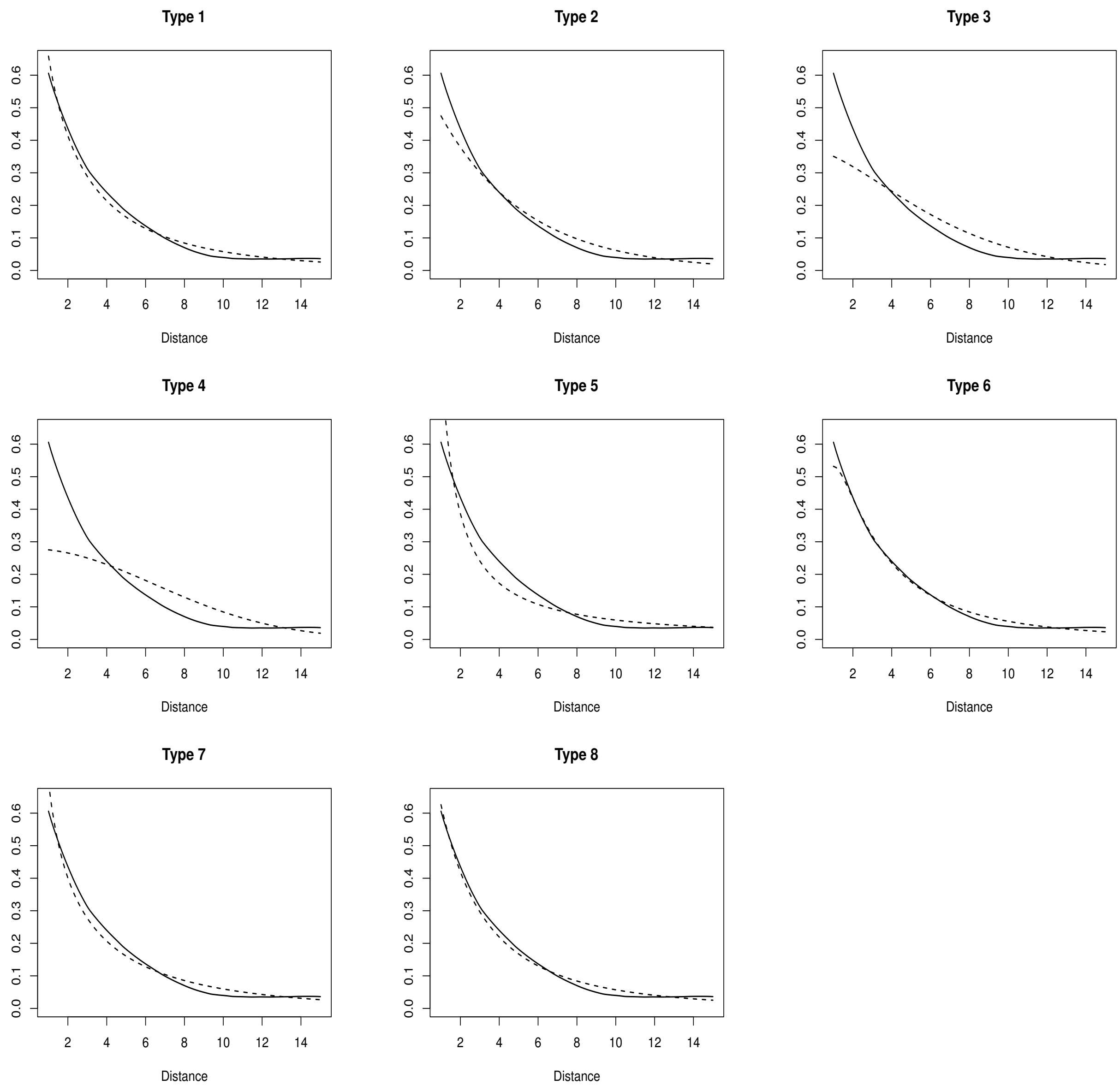

Figure 1: Parametric decay (the dashed line) against the non-parametric decay (the solid line).

patterns. If one is not willing to commit to any parametric decay, the non-parametrically estimated decay can also be used directly.

4 Incorporating Attractiveness

In this section we focus on equation (6) that allows facility attractiveness to interact with the distance decay. The distance decay function depends on the facility attractiveness. More attractive facilities have a slower decay function as was shown in Drezner et al. (2020). Re-stating (6),

$$f(d_{ij}, A_j) = e^{g_0(d_{ij})+g_1(d_{ij})\ln(A_j)}. \tag{19}$$

Within the framework of (19), the parametric distance decays shown in Table 1 can be extended to incorporate the attractiveness level A. The extensions are given in column 2 of Table 2. Other plausible specifications can be similarly specified. For example, to obtain Type 2 parametric specification, one should assume $g_0(d) = \lambda_1 d$ and $g_1(d) = \alpha_1 + \beta_1 d$. If $\beta_1 = 0$, the function in (19) will become the usual exponential decay model $f(d, A) = A^{\alpha_1} e^{\lambda_1 d}$ with $\lambda_1 \leq 0$. Similarly, to obtain Type 5 pattern, one can take $g_0(d) = \lambda_5 \ln d$ and $g_1(d) = \alpha_5 + \beta_5 \ln d$. For $\beta_5 = 0$, this further reduces to the commonly used power decay model $f(d, A) = A^{\alpha_5} d^{\lambda_5}$ with $\lambda_5 \leq 0$. Type 7 pattern is based on $g_0(d) = \lambda_7 d + \lambda_8 \ln d$ and $g_1(d) = \alpha_7 + \beta_7 d + \beta_8 \ln d$.

Estimation of the models in Table 2 mimics the cases of pure distance decay functions that do not depend on the attractiveness level, and can be done within the GLM framework. Replacing $f(d_{ij})$ in (7) by $f(d_{ij}, A_j)$ in (19) and following the same steps, equation (11) is extended to

$$\ln(\mu_{ij}) = \ln(B_i) + g_0(d_{ij}) + g_1(d_{ij})A_j. \tag{20}$$

For model Types 1 - 6, this in turn becomes

$$\ln(\mu_{ij}) = \ln(B_i) + c + \alpha X_{1,ij} + \lambda X_{2,ij} + \beta X_{3,ij} \tag{21}$$

where $X_{1,ij} = \ln A_j$ and $X_{3,ij} = X_{1,ij} \times X_{2,ij}$. For Type 1, $X_{2,ij} = d_{ij}^{0.5}$ and $X_{2,ij} = [\ln(d_{ij})]^2$ for Type 6, and so on. For Types 7 and 8 we have

$$\ln(\mu_{ij}) = \ln(s_i) + c + \alpha X_{1,ij} + \lambda_1 X_{2,ij} + \lambda_2 X_{3,ij} + \beta_1 X_{4,ij} + \beta_2 X_{5,ij} \tag{22}$$

where $X_{4,ij} = X_{1,ij} \times X_{2,ij}$ and $X_{5,ij} = X_{1,ij} \times X_{3,ij}$. For example, for Type 7, $X_{2,ij} = d_{ij}$ and $X_{3,ij} = \ln(d_{ij})$. Because both equations (21) and (22) are linear in their respective parameters, the estimation does not pose any additional complexity except for the addition of more variables due to the interaction terms.

Table 2: Parametric Distance Decays with Attractiveness

Type	$g(d,A)$	$\hat{g}(d,A)$	$\hat{c}$	R^2
1	$\alpha_1 \ln(A) + (\beta_1 \ln(A) - \lambda_1)d^{0.5}$	$-0.06\ln(A) + (0.44\ln(A) - 1.28)d^{0.5}$	-6.17	69.0%
2	$\alpha_2 \ln(A) + (\beta_2 \ln(A) - \lambda_2)d$	$0.21\ln(A) + (0.13\ln(A) - 0.28)d$	-7.44	67.7%
3	$\alpha_3 \ln(A) + (\beta_3 \ln(A) - \lambda_3)d^{1.5}$	$0.31\ln(A) + (0.04\ln(A) - 0.07)d^{1.5}$	-7.92	64.2%
4	$\alpha_4 \ln(A) + (\beta_4 \ln(A) - \lambda_4)d^2$	$0.38\ln(A) + (0.01\ln(A) - 0.02)d^2$	-8.20	59.9%
5	$\alpha_5 \ln(A) + (\beta_5 \ln(A) - \lambda_5)\ln(d)$	$0.44\ln(A) + (0.33\ln(A) - 1.25)\ln(d)$	-7.23	66.6%
6	$\alpha_6 \ln(A) + (\beta_6 \ln(A) - \lambda_6)\ln^2(d)$	$0.40\ln(A) + (0.19\ln(A) - 0.50)\ln^2(d)$	-7.66	69.2%
7	$\alpha_7 \ln(A) + (\beta_7 \ln(A) - \lambda_7)d$ $+(\beta_8 \ln(A) - \lambda_8)\ln(d)$	$0.65\ln(A) + (0.29\ln(A) - 0.27)d$ $-(0.91\ln(A) + 0.10)\ln(d)$	-7.40	70.3%
8	$\alpha_8 \ln(A) + (\beta_9 \ln(A) - \lambda_9)d^{0.5}$ $+(\beta_{10} \ln(A) - \lambda_{10})\ln^2(d)$	$4.73\ln(A) + (-3.57\ln(A) + 1.04)d^{0.5}$ $+(1.54\ln(A) - 0.90)\ln^2(d)$	-8.92	70.4%

4.1 Non-parametric $g_0(d)$ and $g_1(d)$

We suggest a local weighted GLM to estimate $f(d,A)$ in equation (19) by-passing the need to impose parametric functions such as those in Table 2. Assuming that both functions $g_0(d_{ij})$ and $g_1(d_{ij})$ are smooth, it follows from Taylor expansion that

$$\begin{aligned} g_0(d_{ij}) + g_1(d_{ij})\ln(A_j) &\approx g_0(d) + g_0'(d)(d_{ij} - d) + (g_1(d) + g_1'(d)(d_{ij} - d))\ln(A_j) \\ &= \theta_0 + \theta_1 Z_{ij} + \theta_2 X_{1,ij} + \theta_3 X_{2,ij} \end{aligned} \quad (23)$$

where $Z_{ij} = d_{ij} - d$, $X_{1,ij} = \ln(A_j)$ and $X_{2,ij} = Z_{ij} \times X_{1,ij}$. Extending (15) to allow interactions,

$$\ln(E(y_{ij}|d_{ij} = d)) = \ln(s_i) + \theta_0 + \theta_1 Z_{ij} + \theta_2 X_{1,ij} + \theta_3 X_{2,ij} \quad (24)$$

which is locally (at a given d) linear in the parameters. We can adopt the local maximum likelihood steps discussed in Section 3.1.1 where $\theta_0 + \theta_1 Z$ is replaced by $\theta_0 + \theta_1 Z + \theta_2 X_1 + \theta_3 X_2$. The non-parametrically estimated $f(d,A)$ by (19) is given by

$$\hat{f}(d,A) = e^{\hat{\theta}_0(d) + \hat{\theta}_2(d)A} = e^{\hat{g}_0(d) + \hat{g}_1(d)A}. \quad (25)$$

4.2 Comparison of Parametric Specifications

For the seven Orange County, CA shopping malls in our illustration, Drezner (2006) measured their attractiveness ($A_j, j = 1,\ldots,7$) as follows: (South Coast Plaza mall), $A_1 = 2.484$, (Brea mall),

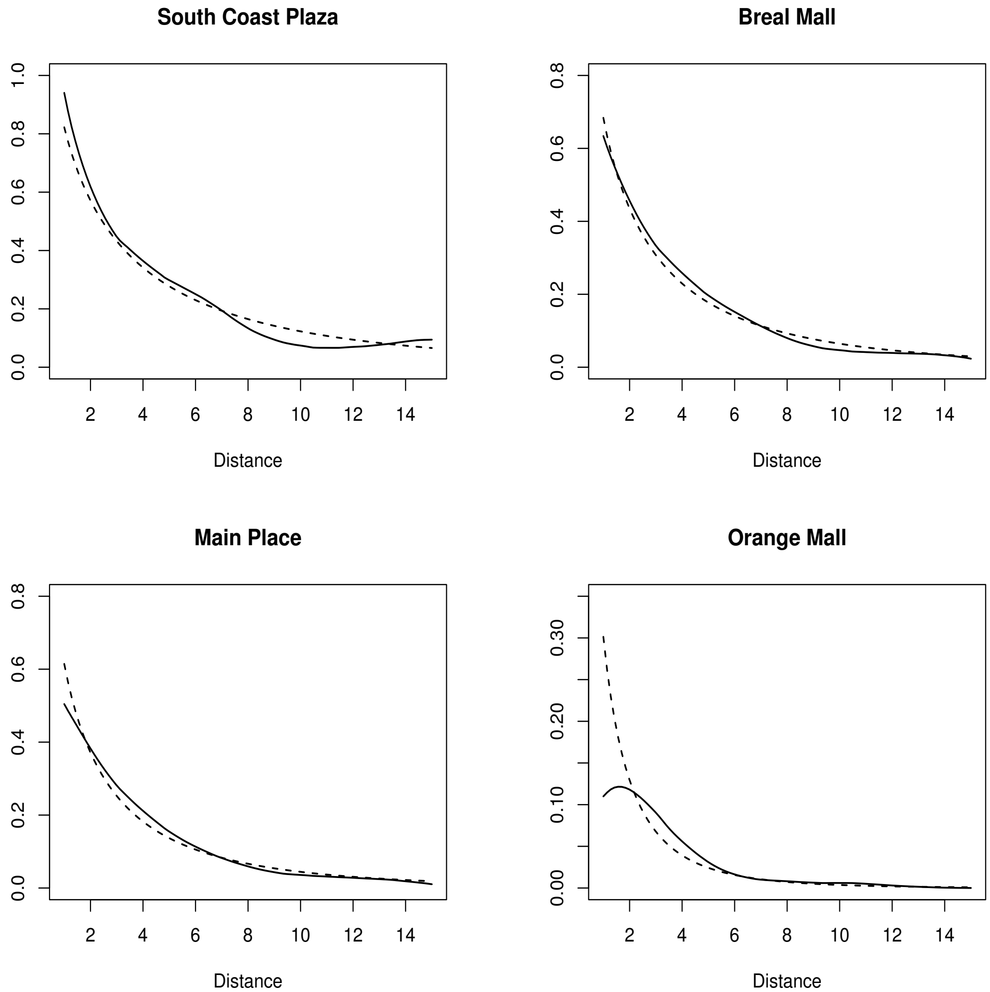

Figure 2: Type 1 parametric model (the dashed line) against non-parametric (the solid line) for four shopping malls.

$A_2 = 1.529$, (Westminster mall), $A_3 = 1.011$, (Main Place mall), $A_4 = 1.154$, (Laguna Hills mall), $A_5 = 0.595$, (Fashion Island mall), $A_6 = 2.367$ and (Orange mall), $A_7 = 0.177$. Extending the pure distance models to allow the attractiveness attribute of shopping malls, Table 2 gives the estimated results for eight parametric models. It is clear that regardless of the form of the parametric specification, attractiveness adds substantial prediction value as reflected in much improved R^2 values. In terms of relative performance, the ranking of the models roughly mimics that for the pure distance cases with Types 1, 6 and 7 and 8 offering better fits.

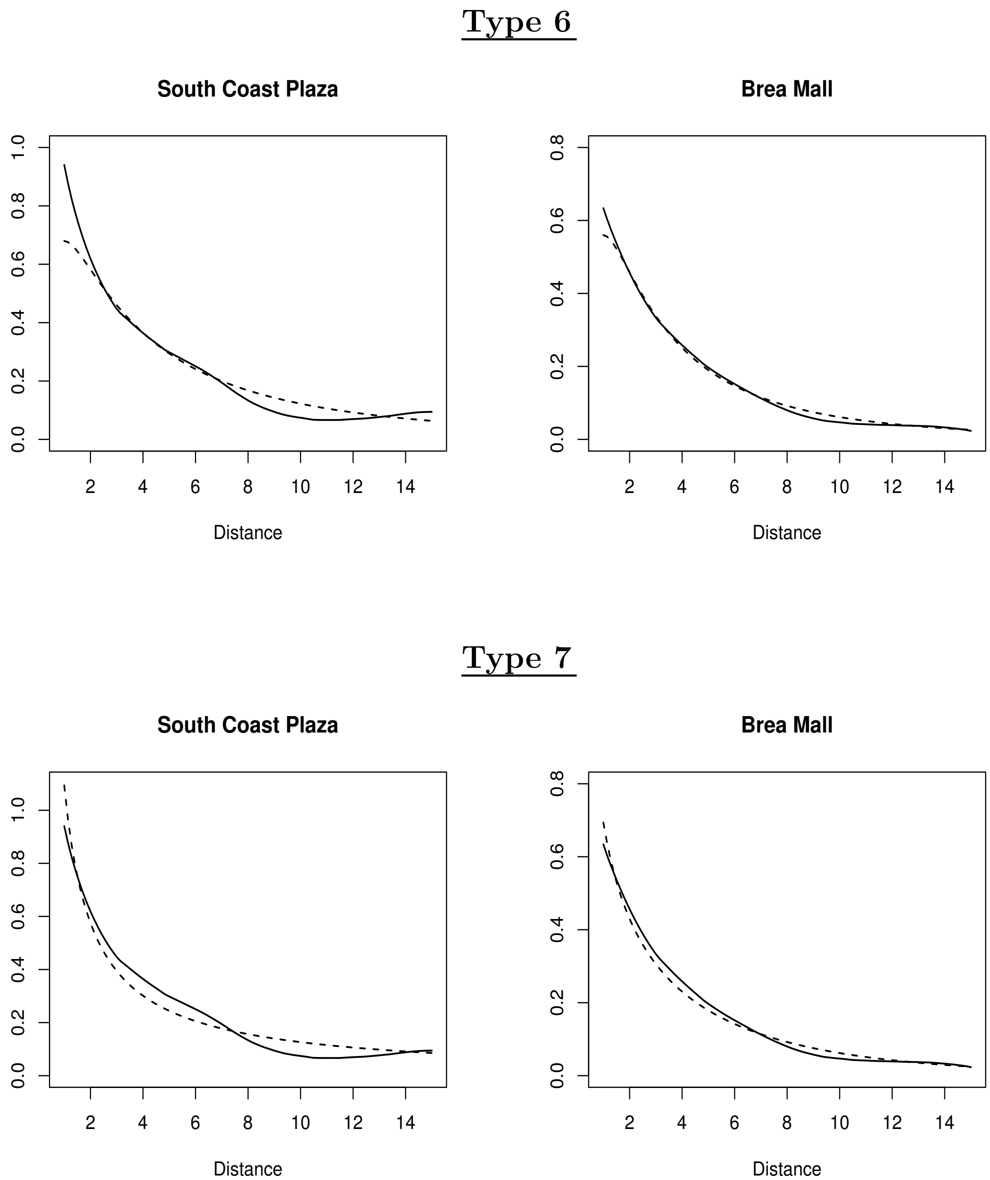

Figure 3: Type 6 and Type 7 parametric model (the dashed line) against non-parametric (the solid line) for two shopping malls.

Because there is a total of 56 cases (8 models and 7 attractiveness levels), we only present selected visualizations that are representative of the overall performances. Accordingly, in Figures 2 and 3, selected estimated distance decays (conditional on a given attractiveness level A_j) for the parametric models are plotted against the non-parametrically estimated decay by equation (25). For example, as shown in Figure 2, Type 1 parametric model closely resembles the non-parametric estimate although its approximation for the least attractive mall (Orange Mall) seems to be off at shorter distances. In Figure 3, visualizations from Type 6 and Type 7 models are shown for two selected malls. The performances relative to the non-parametric pattern are comparable with a slight edge for Type 7 which is consistent with the reported R^2.

5 Locating a New Shopping Mall

Once the distance decay function is selected (we proposed 16 Types, 8 pure distance decay and 8 distance decay with attractiveness), the optimal location for a new facility and the estimated captured market share can be found within a given $\epsilon > 0$ of optimality by global optimization methods. Since the models with attractiveness predict the market share better, we report only the results for the eight models with attractiveness. In Table 3 the expressions for the captured market share based on Table 2 are depicted. We applied the expressions of $\hat{f}(d, \hat{\alpha}, \hat{\beta}, \hat{\lambda})$ for the estimation of the market share.

Table 3: Gravity Models with Attractiveness

Type	$f(d,\alpha,\beta,\lambda)$	$\hat{f}(d,\hat{\alpha},\hat{\beta},\hat{\lambda})$
1	$A^{(\beta\sqrt{d}+\alpha)}e^{-\lambda\sqrt{d}}$	$A^{(0.44\sqrt{d}-0.06)}e^{-1.28\sqrt{d}}$
2	$A^{(\beta d+\alpha)}e^{-\lambda d}$	$A^{(0.13d+0.21)}e^{-0.28d}$
3	$A^{(\beta d\sqrt{d}+\alpha)}e^{-\lambda d\sqrt{d}}$	$A^{(0.04d\sqrt{d}+0.31)}e^{-0.07d\sqrt{d}}$
4	$A^{(\beta d^2+\alpha)}e^{-\lambda d^2}$	$A^{(0.01d^2+0.38)}e^{-0.02d^2}$
5	$A^{(\beta \ln d+\alpha)}e^{-\lambda \ln d}$	$A^{(0.33\ln d+0.44)}e^{-1.25\ln d}$
6	$A^{(\beta \ln^2 d+\alpha)}e^{-\lambda \ln^2 d}$	$A^{(0.19\ln^2 d+0.40)}e^{-0.50\ln^2 d}$
7	$A^{(\beta_1 d+\beta_2 \ln d+\alpha)}$ $\times e^{-\lambda_1 d+\lambda_2 \ln d}$	$A^{(0.29d-0.91\ln d+0.65)}$ $\times e^{-0.27d-0.10\ln d}$
8	$A^{(\beta_1\sqrt{d}+\beta_2 \ln^2 d+\alpha)}$ $\times e^{-\lambda_1\sqrt{d}+\lambda_2 \ln^2 d}$	$A^{(-3.57\sqrt{d}+1.54\ln^2 d+4.73)}$ $\times e^{1.04\sqrt{d}-0.90\ln^2 d}$

The common procedures applied for the location of one facility are the Big Square Small Square (BSSS, Hansen et al., 1981) applied, for example, in Drezner and Zerom (2023a), and Big Triangle Small Triangle (BTST, Drezner and Suzuki, 2004) applied, for example, in Drezner and Drezner (2004). A general approach for generating the upper bound is proposed in Drezner (2007). We applied the BSSS algorithm because the data is given in a square. Run times are negligible (a very small fraction of a second) and thus not reported.

The expected market share captured by a facility from a demand point decreases as a function of the distance in all competitive location models, including, of course, the gravity model and the 16 distance decay types tested in this paper. Therefore, a simple upper bound for a facility located anywhere in a square or a triangle is the calculated market share at the closest possible distance between a demand point and the square or the triangle. The formula for the shortest distance to a square is given in (Drezner et al., 2023; Drezner and Zerom, 2023a), and the shortest distance to a triangle is given in Drezner and Drezner (2004).

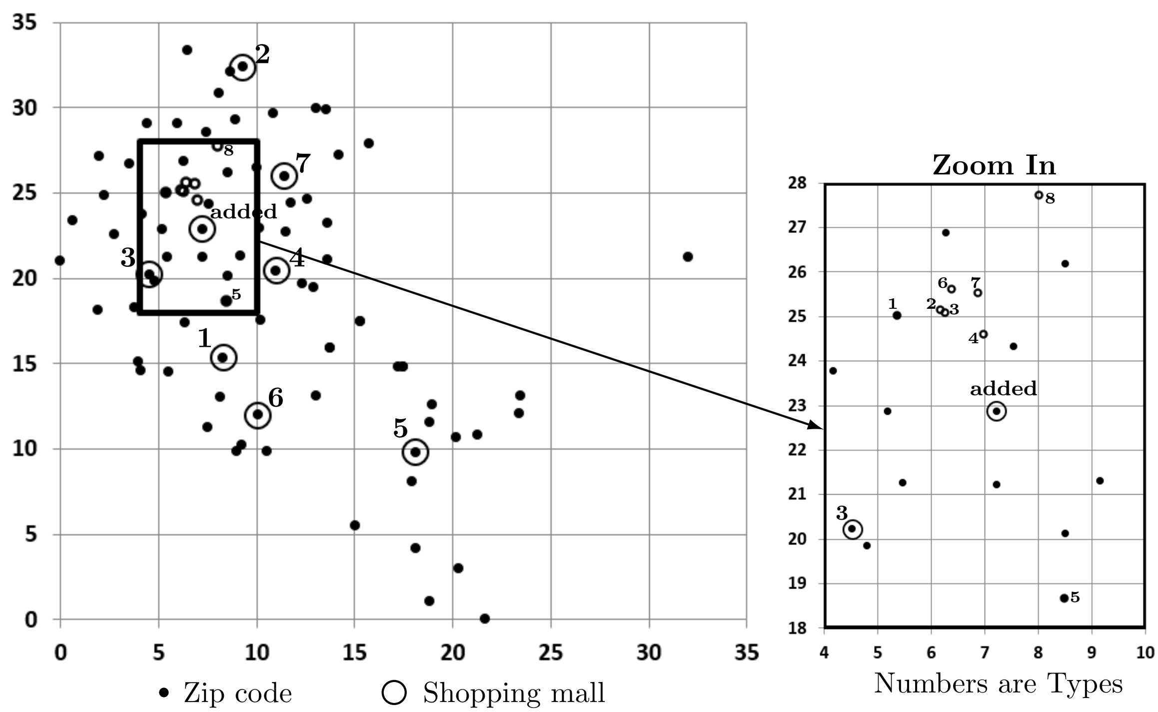

Figure 4: Zip codes, mall locations, and locations of a new facility

We found the optimal location and captured market share for a low-attractive new facility ($A = 0.2$), average attractive new facility ($A = 1$), and a highly attractive new facility ($A = 2$). The locations of 80 zip codes and the seven malls, in addition of some results are depicted in

Figure 4. The coordinates and properties of the shopping malls are available in Drezner et al. (2020). The best locations for a new mall, as also found in Drezner et al. (2020), are in the area between facilities 3, 4, and 7 which is quite densely populated. There is a large strip plaza located in this area. Also, north of Orange County is Los-Angeles County that has malls that serve some of the population in that area. Our data set does not include these competing facilities, so the best location for a new mall is expected to be in that area when the tested data set is applied. We therefore added an eight facility with $A = 2$ at the best location for Type 8 which is depicted in Figure 4. We "assume" that when a decision is made to locate a new facility it is known that a facility is under construction at that location.

Table 4: Locations of a New Facility for Parametric Decay with Attractiveness

	$A = 0.2$			$A = 1$			$A = 2$		
Type	x	y	MS	x	y	MS	x	y	MS
1	5.37	25.02	4.295%	5.37	25.02	9.913%	8.51	20.11	16.386%
2	6.18	25.15	3.535%	7.35	23.29	9.871%	8.21	22.11	16.721%
3	6.26	25.08	3.497%	7.37	23.23	9.820%	8.24	22.08	16.661%
4	6.98	24.59	4.164%	7.57	23.51	10.607%	8.24	22.62	16.450%
5	8.50	18.67	4.919%	8.50	18.66	10.315%	8.51	20.11	16.413%
6	6.38	25.61	3.697%	6.55	23.87	10.702%	7.89	21.91	18.295%
7	6.88	25.52	3.334%	5.37	25.02	9.892%	8.51	20.11	16.807%
8	8.02	27.71	3.650%	6.86	25.33	9.807%	8.51	20.11	18.306%

Table 5: Decline in Market Share Captured

Type	$A = 0.2$	$A = 1$	$A = 2$
1	-6.47%	-10.94%	0%
2	-6.06%	-0.76%	-9.21%
3	-5.65%	-0.74%	-9.17%
4	-2.96%	-1.48%	-9.92%
5	-26.25%	-17.94%	0%
6	-3.84%	-0.16%	-8.81%
7	-1.15%	-10.94%	0%
8	0%	0%	0%

The locations and market shares captured, when another competing facility is added to the list of existing facilities, are listed in Table 4. For $A = 2$ the locations are in a small area for all eight types. However for $A = 0.2$ and $A = 1$ the location for Type 5 is changed significantly. It is strange that the captured market share for $A = 0.2$ is largest for Type 5. This is explained by the fact that the expected market share for Type 5 is not accurate. The most accurate estimate is obtained by Type 8 that has the largest R^2 value as reported in Table 2. We therefore calculated the market share captured at the optimal location for each type by the Type 8 formula. In Table 5 the loss in market share at the location for a certain type compared with the more accurate evaluation of the market share by the Type 8 formula is reported. For example, for $A = 0.2$ the market share captured at the Type 5 location is 2.691% which is 26.25% below the optimal market share of 3.650% reported in Table 4 for Type 8. A user that applies the power decay (Type 5) would locate at a much inferior location. He will expect a market share of 4.919% and will be "surprised" to find out after the facility was located, that he captures only 2.691% which is only 55% of what he expected.

In Figure 4 the eight locations for the facilities of attractiveness 0.2 are depicted. The area where all new facilities are located is zoomed in on the right so that the eight locations can be clearly seen. In spite of adding the new eighth facility, many of the locations are close to the area between facilities 3, 4, and 7 but most are located more to the north. The Type 5 location is outside the area further to the south, and the Type 8 location, which is the recommended location by the R^2 criterion, is located more to the north closer to Brea Mall (facility #2).

6 Conclusions

Competitive facilities location problems attempt to find the locations for one or more new facilities among existing competing facilities that maximize the captured market share. The facilities attract demand generated by customers in the area. Many approaches to estimate the captured market share were proposed over the last 100 years. The gravity model (Reilly, 1931) is considered to be the most accurate one. We introduce a data-driven gravity model where no assumption is made about the functional form of the distance decay.

The approach is unifying in the sense that it encompasses practically all existing parametric approaches (such as the power and exponential decays) as special cases. To facilitate practical implementation, we also provide an estimation procedure that is based on the generalized linear models (GLM) that is suited both for non-parametric and parametric gravity models. To the best of our knowledge, a non-parametric gravity model is new to the literature and is quite timely as data have become easier to collect and process. To illustrate the contribution, we use real data on shopping malls in Orange County, California. Although the scale of the data is rather small, the example is informative to highlight the value of the data-guided approach and how it can also serve to empirically validate existing parametric distance decays.

In the gravity model it is assumed that the probability that a customer patronizes a facility is proportional to the facility's attractiveness and to a distance decay function. Before the paper by Drezner et al. (2020), most competitive location papers assumed that the distance decay function is the same for all facilities, and concentrated on approaches to estimate the attractiveness levels of the facilities which is estimated by public opinion surveys. For example, Nakanishi and Cooper (1974) suggested that the attractiveness is proportional to a product of characteristics and each characteristic is rated by customers surveys. However, estimating the expected market share by actual customer behavior is more accurate than relying on public opinion surveys. We apply the data based on intercepting customers and inquiring about their zip code, which measures their actual behavior, rather than their opinion which is less reliable. A more accurate estimate of the captured market share leads to establishing new competing facilities in better locations by the optimization procedures.

Funding Declaration: There was no funding for this manuscript.

References

Aboolian, R., Berman, O., and Krass, D. (2007a). Competitive facility location and design problem. *European Journal of Operations Research*, 182:40–62.

Aboolian, R., Berman, O., and Krass, D. (2007b). Competitive facility location model with concave demand. *European Journal of Operations Research*, 181:598–619.

Bell, D., Ho, T., and Tang, C. (1998). Determining where to shop: Fixed and variable costs of shopping. *Journal of Marketing Research*, 35(3):352–369.

Berman, O., Krass, D., and Wang, J. (2006). Locating service facilities to reduce lost demand. *IIE Transactions*, 38:933–946.

Downs, R. M. (1970). The cognitive structure of an urban shopping center. *Environment and Behavior*, 2:13–39.

Drezner, T. (1994). Optimal continuous location of a retail facility, facility attractiveness, and market share: An interactive model. *Journal of Retailing*, 70:49–64.

Drezner, T. (2006). Derived attractiveness of shopping malls. *IMA Journal of Management Mathematics*, 17:349–358.

Drezner, T. (2023). Stochastic components of the attraction function in competitive facilities location. In *Uncertainty in Facility Location Problems*, pages 107–127. Springer.

Drezner, T. and Drezner, Z. (2004). Finding the optimal solution to the Huff competitive location model. *Computational Management Science*, 1:193–208.

Drezner, T. and Drezner, Z. (2008). Lost demand in a competitive environment. *Journal of the Operational Research Society*, 59:362–371.

Drezner, T. and Drezner, Z. (2012). Modelling lost demand in competitive facility location. *Journal of the Operational Research Society*, 63:201–206.

Drezner, T., Drezner, Z., and Zerom, D. (2020). Facility dependent distance decay in competitive location. *Networks and Spatial Economics*, 20:915–934.

Drezner, T., Drezner, Z., and Zerom, D. (2022). An extension of the gravity model. *Journal of the Operational Research Society*, 73:2732–2740.

Drezner, T., O'Kelly, M., and Drezner, Z. (2023). Multipurpose shopping trips and location. *Annals of Operations Research*, 321:191–208.

Drezner, Z. (2007). A general global optimization approach for solving location problems in the plane. *Journal of Global Optimization*, 37:305–319.

Drezner, Z. and Eiselt, H. (2023). Competitive location models: A review. *European Journal of Operational Research*. DOI: 10.1016/j.ejor.2023.10.030.

Drezner, Z. and Suzuki, A. (2004). The big triangle small triangle method for the solution of non-convex facility location problems. *Operations Research*, 52:128–135.

Drezner, Z. and Zerom, D. (2023a). Competitive facility location under attrition. *Computational Management Science*. DOI:10.1007/s10287-023-00473-z.

Drezner, Z. and Zerom, D. (2023b). A refinement of the gravity model for competitive facility location. *Computational Management Science*. DOI: 10.1007/s10287-023-00484-w.

Eiselt, H. A., Marianov, V., and Drezner, T. (2015). Competitive location models. In Laporte, G., Nickel, S., and da Gama, F. S., editors, *Location Science*, pages 365–398. Springer.

Fernández, J., Pelegrin, B., Plastria, F., and Toth, B. (2007). Solving a Huff-like competitive location and design model for profit maximization in the plane. *European Journal of Operational Research*, 179:1274–1287.

Hansen, P., Peeters, D., and Thisse, J.-F. (1981). On the location of an obnoxious facility. *Sistemi Urbani*, 3:299–317.

Hodgson, M. J. (1981). A location-allocation model maximizing consumers' welfare. *Regional Studies*, 15:493–506.

Huff, D. L. (1964). Defining and estimating a trade area. *Journal of Marketing*, 28:34–38.

Huff, D. L. (1966). A programmed solution for approximating an optimum retail location. *Land Economics*, 42:293–303.

Jain, A. K. and Mahajan, V. (1979). Evaluating the competitive environment in retailing using multiplicative competitive interactive models. In Sheth, J. N., editor, *Research in Marketing, Vol. 2*, pages 217–235. JAI Press, Greenwich, CT.

McCullagh, P. and Nelder, J. A. (2019). *Generalized linear models*. Routledge.

Nakanishi, M. and Cooper, L. G. (1974). Parameter estimate for multiplicative interactive choice model: Least squares approach. *Journal of Marketing Research*, 11:303–311.

Prosperi, D. C. and Schuler, H. J. (1976). An alternate method to identify rules of spatial choice. *Geographical Perspectives*, 38:33–38.

Reilly, W. J. (1931). *The Law of Retail Gravitation*. Knickerbocker Press, New York, NY.

Schuler, H. J. (1981). Grocery shopping choices: Individual preferences based on store attractiveness and distance. *Environment and Behavior*, 13:331–347.

Timmermans, H. (1982). Consumer choice of shopping centre: an information integration approach. *Regional Studies*, 16:171–182.

Timmermans, H. (1988). Multipurpose trips and individual choice behaviour: an analysis using experimental design data. In *Behavioural modelling in geography and planning*, pages 356–367. Croom Helm.

Wand, M. P. and Jones, M. C. (1994). *Kernel smoothing*. CRC press.

Wilson, A. G. (1976). Retailers' profits and consumers' welfare in a spatial interaction shopping mode. In Masser, I., editor, *Theory and Practice in Regional Science*, pages 42–59. Pion, London.