

Analyzing the relationship between stadium location and attendance: A location modeling of Major League Baseball (MLB)

Namhun Lim^{1*}, Wanyong Choi², Paul Mark Pedersen³

¹*Department of Sport Management, Wichita State University, USA*

²*Department of Leadership, School Counseling, & Sport Management, University of North Florida, USA*

³*Department of Kinesiology, Indiana University Bloomington, USA*

Abstract

The location of a sports facility is one of the significant factors in economically operating and maintaining the facility. The present study used a location theory model to examine the influence of the Major League Baseball (MLB) stadium location on attendance (N= 348) for the 2006-2017 MLB regular seasons, taking into consideration the population, median income, and distance between the stadium and the population center of counties nearby the stadium. The optimal location model was determined as the inverse proportion of the squared distance of the first nearest county, and location significantly influenced the attendance ($p < .05$). Then, based on the results of the location model, the current study analyzed the relationship between the MLB attendance and other attendance determinants by using hierarchical linear modeling. Among 14 different determinants, six variables (i.e., stadium capacity, ticket price, season, as well as the home team's quality, payroll, and star players) were revealed in the findings to have a significant influence on the MLB attendance ($p < .05$).

Key words: Stadium, Location, Attendance, MLB stadium, Location modeling, Hierarchical linear modeling

Introduction

In 2016, residents of Arlington, Texas voted in favor of an increase in the sales tax on hotel stays and rental cars in order to fund a new stadium for the city's Major League Baseball (MLB) team, the Texas Rangers, even though Rangers had used the current home stadium, Global Life Park, for only 23 years (Perry, 2016).

As the mayor of Arlington, Jeff Williams, argued in favor of the economic impact that would be realized with the arrival of a new stadium (Formby, 2016), some studies (e.g., Gratton, Shibli, & Coleman, 2005; Perry, 2001; Tu, 2005) have shown the benefits of investing in new stadiums or hosting sporting events because such endeavors can at times lead to an increase in attendance and in the number of tourists, new employment opportunities, image enhancement of the city, and an increase in housing values. Other studies, however, have found that the expectations of such tangible economic benefits through investments in stadiums are often not realized and that involvement in such eco-

conomic expenditures is not a proper activity for public agencies (e.g., Baade, 1996; Zimbalist, 1998). Based on analyses of positive economic benefits (e.g., enhancing the image of the city or country, new employment opportunities), some politicians (e.g., Mayor Williams) or sport organizations have attempted to persuade and promote to the public the advantages of their proposed investments in new stadiums or in hosting hall-mark sporting events (e.g., Whitson & Horne, 2006). However, it is necessary to analyze the actual expected gains, and to deliberate before deciding, because the major portion of the investment in the new facility will come from public sources such as general taxes, selective taxes (e.g., tourist taxes, sin taxes, players' income taxes, ticket surcharges), and debt financing (Howard & Crompton, 2014).

Indeed, the cost of building new stadiums for sporting events or sports leagues receives the most media attention because construction costs are spent in the early stages of stadium development. However, because the high costs are invested, the operation of the new sport facilities should be fully considered to cover the costs invested or to prevent any additional (future) investments. Thus, deciding on the location of a stadium in a city is important for both the public and for managers of sporting events or sports teams because costs (e.g., maintenance expenses) and revenues (e.g., gate receipts, sponsorship) often are dependent on the location of the stadium. For example, in 2012 the Miami Marlins moved from Sun Life Stadium to the newly constructed Marlins Park. One of the reasons that the Marlins moved to the new stadium had to do with the obvious advantages of a domed stadium in Florida's often-unpredictable weather (Kimmelman, 2012). While the Marlins' move was due to a search for a better environment for both the players and the spectators, enhanced accessibility was a key motivation behind the recent move by the Atlanta Braves from Turner Field to the newly built SunTrust Park (Jaffe, 2013). The Braves expect higher attendance numbers with their new stadium's location, which is now situated in the

geographic center of the Braves' fan base (Tucker, 2013). While stadium location is important in new construction decisions, most of the studies involving MLB attendance are focused on attendance determinants rather than stadium location. Thus, the current study fills this gap in the literature by investigating the effect of stadium location on attendance in MLB.

In the case of the Atlanta Braves' relocation to their new home stadium, it is clear that the decision regarding the location of the stadium was made after consideration of various economic and demographic aspects. For this reason, numerous studies have investigated the significant effects of economic and demographic factors, such as population and household income levels on attendance at sporting events (e.g., Coates & Humphreys, 2007; Hansen & Gauthier, 1989; McEvoy, Nagel, DeSchrive, & Brown, 2005). However, distance is considered an indicator of other options that might be available for viewing other games (Winfree, McCluskey, Mittelhammer, & Fort, 2004) and rivalry games (DeSchrive & Jensen, 2002; Lemke, Leonard, & Tlhokwane, 2010). While the distance factor in the study by Winfree and colleagues was measured in miles to the closest stadium, the studies by DeSchrive and Jensen, and Lemke et al. used the distances between the home team and the visiting team. However, few previous studies (e.g., Nelson, 2002) have been concerned with stadium location. Although the study of Nelson investigated economic activities by stadium location at downtown, edge, and suburban areas, it did not involve a study of the attendance demand. Therefore, the purpose of this study was to investigate the influence of stadium location on attendance, utilizing the well-known determinants found in other studies (e.g., DeSchrive & Jensen, 2002; Rivers & DeSchrive, 2002) of attendance in MLB.

Attendance Determinants in Sporting Events

Howard and Crompton (2014) emphasized and confirmed that ticket sales of sporting events have been

considered as a primary source of revenue. Coming in the form of ticket sales or gate receipts, numerous studies (e.g., Hansen & Gauthier, 1989; McDonald & Rascher, 2000) have illustrated the importance of attendance demand in sporting events. Among the various studies of attendance determinants in sporting events, of particular importance are the five categories of Borland and Macdonald (2003) which emphasized the viewing quality of sporting events and the four categories of Schofield (1983) which focused more on demographic and economic factors. The present study employed Schofield's classification, which included game attractiveness factors, economic factors, demographic factors, and residual preference factors. The rationale behind this selection was that the current study focused more on the effects of demographic and economic factors of attendance. Various studies (e.g., Ferreira & Bravo, 2007; Hansen & Gauthier, 1989) have examined determinants based on Schofield's classification.

Attractiveness factors

Previous studies (e.g., Carmichael, Millington, & Simmons, 1999; Hansen & Gauthier, 1989) have noted that there are attractiveness factors that affect attendance. These attraction determinants are related to the quality of sports content such as the win-loss records of the home and visiting teams (e.g., Carmichael et al., 1999; Hansen & Gauthier, 1989); the number of star players (Rivers & DeSchrive, 2002); the rankings of the home and visiting teams in the league and division standings (e.g., Siegfried & Eisenberg, 1980); the chance of a home or visiting team advancing to the playoffs and rivalry games between home and visiting teams (Lemke et al., 2010); the offensive and defensive performance of each team (Lemke et al., 2010; McDonald & Rascher, 2000; Siegfried & Eisenberg, 1980); the past successes of home and visiting teams (Ferreira & Bravo, 2007; Lemke et al., 2010); and special promotions, events, or occasions (Hansen & Gauthier,

1989). The attractiveness factors (e.g., team qualities, match uncertainty, season ranks, team performances, star players) would be related to the internal factor of sporting events (e.g., game on field).

Economic factors

The economic factors such as ticket price and income level of local population influence the sporting events attendance because attending a sporting event is a leisure activity that heavily depends on consumption behavior. Among various economic indicators, Hansen and Gauthier (1989) introduced eight variables in the categories of economic factors: television coverage of the home game in the local area; price of season tickets for home games; television coverage of another major sporting event at the same time as a home game; price of the ticket for a home game; price of other forms of entertainment available during a team's home games; existence of other sporting teams in the area; average income of the population; and other professional franchises in the area. These economic factors categories are related to purchasing power, and to the level of substitute activities (e.g., other sporting teams, other league games) within the local area.

Demographic factors

In addition to economic factors, demographic factors also influence the attendance at sporting events. In the category of demographic factors, Hansen and Gauthier (1989) described three variables as the population of city of home team, the ethnic diversity of the population, and the existence of minor league sports for children and youth. Generally, market size is directly proportional to the size of the population. Because this notion is also applicable to the sport industry, most professional teams are located in cities with a large population. Due to the importance of population in any examination of attendance demand, several studies (Coates & Humphreys, 2007; Ferreira & Bravo, 2007;

Lemke et al., 2010; McDonald & Rascher, 2000; McEvoy et al., 2005; Siegfried & Eisenberg, 1980) have included a population variable as one of the predictors of attendance.

The popularity of sports differs, depending upon race and ethnicity. That is, specific sports are favored by specific ethnic group, the ethnic diversity of in a team's home city is a significant determinant of attendance at sporting events (e.g., Siegfried & Eisenberg, 1980). Therefore, these demographic factors directly affect the total sales of business including gate receipts because demographics (e.g., population of city, the ethnic diversity of the population) indicate the market size in the area (Campbell & Hopenhayn, 2005). In addition, the existence of minor league sports for children and youth serves as a determinant of attendance. The experience of youth sport participation is regarded as a significant determinant that affects sport activity involvement in adulthood (Vanreusel et al., 1997). Numerous other studies have supported the notion that participation in physical activity in childhood influences a person's life into adolescence and adult (e.g., Kemper, De Vente, Van Mechelen, & Twisk, 2001).

Residual preference factors

The residual preference factors are related to the quality of viewing for the spectators. The residual preference category factors include the quality of the stadium, the convenience and ease of access to the facility, and the day and time of the game (Hansen & Gauthier, 1989). The attractiveness factors are associated with the preferences of attendees for the external game elements such as weather (e.g., temperature) (e.g., McDonald & Rascher, 2000), game day and time (e.g., Carmichael et al., 1999; Lemke et al., 2010), the convenience of transportation (Hansen & Gauthier, 1989), and the stadium's comfort for the spectators (Borland & Macdonald, 2003; Ferreira & Bravo, 2007).

Based on the Schofield's (1983) four categories of attendance determinants, various elements including

demographic (e.g., population, ethnic mix of the population) and economic factors (e.g., income level, ticket price, television coverage), game attractiveness (e.g., team qualities, match uncertainty, team performances, star players), and residual preference factors (e.g., age and capacity of stadium, day of week and time of game) have been investigated with regard to their effect on attendance at sporting events (e.g., Coates & Humphreys, 2007). However, few studies have examined stadium location as one of the attendance determinants even though optimal location is an important factor that benefits facility operations (Snyder & Daskin, 2005). Therefore, it is necessary to understand how to apply location modeling to sport facility locations.

Location Modeling in Sports

Various determinants (e.g., economic conditions, demographics, game attractiveness, residual preference factors) influence the attendance in sporting events. Though few studies point to location factors as attendance determinants in the sport industry, numerous researchers over a long period of time have been interested in developing various forms of optimal location models for such facilities as factories, warehouses, shopping malls, schools, hospitals, and others (e.g., Aghezzaf, 2005; Cheng, Li, & Yu, 2005). Although the objectives or goals of each of these location studies are different, most have included common main features, such as examinations of the sizes, metrics, customers, and facilities (ReVelle & Eiselt, 2005; ReVelle, Eiselt, & Daskin, 2008). In addition, scholars (e.g., Hamacher & Nickel, 1998; ReVelle et al., 2008) have identified four broad categories of location modeling (i.e., analytic models, continuous models, network models, and discrete location models) even though various classifications of location modeling have existed.

The first of the four previously identified categories – which is referred to as analytic models – is based on simplified assumptions of real situations. The typ-

ical assumptions of analytic models are that demand, with constant density, is equally distributed over a service area, and a fixed cost exists for operating a facility. Also, the unit shipment cost per distance is constant no matter where the facility is located. Based on these assumptions, the total costs of operating facilities can be expressed using terms of these fixed and operating costs. It is also possible to analyze the optimal total costs over a number of facilities (Leamer, 1968; ReVelle et al., 2008). However, the limitation of the analytic models is that there is a difference between the assumptions of the simplified model and the factors found in real situations (ReVelle et al., 2008).

The second category of location models is known as continuous models. Among all of the location models, continuous models have been used the longest (Hamacher & Nickel, 1998). Also, ReVelle et al. (2008) described that there are two central suppositions of continuous models. The first is that facilities could be in any service area, and the second is that separate demands exist at each discrete location. A single facility could also meet several demands, with coordinated information and weighted demand. Distance is measured via a straight line between the facility and each demand, and thus the optimal location of any single facility would be at the location of minimized total demand-weighted distance. Continuous models also assume that a single facility may be located at any place within any service area. However, because any facility needs a specific environment, this assumption is limited.

In addition, ReVelle et al. (2008) explained both the third and fourth categories of location models: network models and discrete location model. Network models assume the existence of a network that is composed of links and nodes. Demands exist at each node, and links play the role of connecting the nodes. Generally, network model studies center around graph structures that specify location problems and are represented by nodes and links. These nodes and links use simplified structures such as low-order polynomial time algo-

rithms. However, discrete location models assume the existence of a discrete set of demands and proposed locations. While network models are concerned with defining graphical structures, discrete location models focus on finding efficient heuristic algorithms based on various practical cases and experiences. Furthermore, in cases where there are problems in a maximum distance model in discrete location models, the facility considers covering the maximum service area within a specific time frame (Current, Daskin, & Schilling, 2004). Also, typical p -median location models (Hakimi, 1964, 1965) are concerned with finding a location that minimizes the total demand-weighted distance.

Most location models are developed for the purpose of making decisions about the optimal location of a facility. Optimal location leads to an increase in efficiency by covering more of the demand area and minimizing the operational costs. In order to apply this concept to determine the optimal location of a stadium, it is necessary to identify a consumer's desire for attendance and the distance between that demand and the stadium. The primary consumers attending sporting events are people who are interested in sports. Even though the proportion of fans of specific sports among the total population depends on the type of sport, potential spectators are proportional to the size of the population regardless of the type of sport if the general economic concept, that market size is directly proportional to population (Campbell & Hopenhayn, 2005), is applied to sports. Due to the importance of the size of the population, several studies have investigated the relationships between population size and attendance in MLB (e.g., Coates & Humphreys, 2007; Lemke et al., 2010; McDonald & Rascher, 2000), the minor leagues (Siegfried & Eisenberg, 1980), the National Basketball Association (NBA) (Coates & Humphreys, 2007), and the Chilean national soccer tournaments (Ferreira & Bravo, 2007). Moreover, the income level of an area is an indication of its potential purchasing power, and has been found to have positive influences on attendance at sporting events in

various studies (e.g., Hansen & Gauthier, 1989). Based on the results of previous studies (e.g., Coates & Humphreys, 2007), population size and income level have been found to be factors that determine the demands for attendance at sporting events.

The distance between each stadium and the location of each individual consumer is impossible to measure because of the huge number of potential consumers in every city. However, by using a process of demand data aggregation, demand nodes may be reduced (Current & Schilling, 1987). This approach uses the distance between the facility and an aggregated node, and the median distance among all distances between the facility and the individual demand location is determined. This is called the aggregation process, and it has been employed by several location studies in the field of urban services facilities, power plant, fire station, and other applications (e.g., Francis, Lowe, & Tamir, 2004; Serra & Marianov, 1998).

Based on the assumption of continuous models, it may be assumed that potential sport spectators are equally distributed within a city or county. By using this assumption and the demand data aggregation process, the distance between the stadium and individual demands may be found by the distance between the stadium and the mean center of population.

Using the literature outlined above, the present study was constructed to investigate the relationship between the attendance and the stadium location of each MLB team. Based upon the previously mentioned prior studies of attendance demand for sporting events and location theory, the following research questions guided this investigation:

RQ1: What is the relationship between attendance demand and the location of stadiums in Major League Baseball (MLB)?

RQ2: What is the maximum number of counties that are near stadiums that have an influence on attendance in MLB?

RQ3: How does attendance demand correlate with

stadium location and the observed determinants of the four categories (i.e., economic factors, demographic factors, game attractiveness factors, and residual preference factors) in the MLB regular season?

Methodology

To analyze the relationship between the actual MLB attendance and the attendance determinants (e.g., demographics, economics, location factors), this study employed seasonal average attendance ($N = 348$) of 29 MLB teams from the MLB regular seasons ranging from 2006 to 2017 ("MLB Attendance Report," 2017), as the dependent variable. In addition, three independent variables (i.e., county's population, income, and distance) in the location modeling and 14 independent variables in the attendance demand analysis were included and used in this study. Especially, to analyze the relationship between attendance and stadium location (RQ1 and RQ2), economic, demographic, and geographic factors, such as population, income level, and distance between stadium and population center, were adopted and utilized within the location model for presenting market size and purchasing power. While there are challenges involved with finding a true and exact center of population distribution by city, the present study addressed this concern by employing demographic data by county.

The data associated with population sizes and the median household incomes of counties were taken from the U.S. Census Bureau (2018). The distances were measured via a straight line between the stadium and the center of the county population (U.S. Census Bureau, 2010) as seen on Google Maps. However, the Canadian census does not provide information about the regional population center and thus this study excluded the Toronto Blue Jays (which is why the study used data related to 29 instead of all 30 MLB teams). These population sizes, median incomes, and distances were adopted as independent variables to investigate the relationship between attendance and stadium loca-

tion (RQ1). Based on these variables, the location model of an MLB stadium was determined using the following formula:

$$Att_{ij} = \alpha_0 + \alpha_1 \left(\frac{P_{1ij} * I_{1ij}}{D_{1ij}^x} \right) + \epsilon_{ij}$$

Where, Att_{ij} : average attendance of team j in season i

- P_{1ij} : population size of nearest county
- I_{1ij} : median household income of nearest county
- D_{1ij} : distance between stadium and population center of nearest county
- i : season i ($i = 1, \dots, 12$)
- j : team j ($j = 1, \dots, 29$)

From the results of the model fit comparison (i.e., -2 log likelihood, AIC, and BIC) for the optimal various (x) values, the effect of the distance (x) between the stadium and the county on MLB attendance could be determined. Based on the results of this model, a demand-weighted location model was constructed using the effect of the distance (x) with data of not only the nearest county to an MLB team, but also other counties near the stadium. The demand-weighted location model of a stadium was constructed thus:

$$Att_{ij} = \beta_0 + \beta_1 Location_{ij} + \epsilon_{ij}$$

$$= \beta_0 + \beta_1 \left(\frac{P_{1ij} * I_{1ij}}{D_{1ij}^x} + \frac{P_{2ij} * I_{2ij}}{D_{2ij}^x} + \dots + \frac{P_{nij} * I_{nij}}{D_{nij}^x} \right) + \epsilon_{ij}$$

Where, n : county of n^{th} nearest from stadium ($n = 1, \dots, 5$)

In order to investigate if the number of counties near stadiums influences MLB attendance (RQ2), the optimal n is decided via this demand-weighted location model and the model fit tests (i.e., -2 log likelihood, AIC, and BIC).

Through finding the distance effect (x) of (n) counties on attendance, the location variable of MLB stadium was decided. Also, it was possible to construct the full model of seasonal average attendance demand with other attendance determinants that were elimin-

ated for reducing biases at the initial location models. Hierarchical linear modeling (HLM), also known as multilevel modeling (MLM) was also employed to provide an efficient explanation of the differences between seasons and between teams. Table 1 shows the variables that are employed in the model.

After the data were collected, the two-level null model was constructed to investigate the relationship between the attendance during the MLB regular season and the determinants. In order to examine the relationship between stadium location and the observed deter-

Table 1. Variable descriptions of MLB attendance demand

Symbol	Description
Dependent variable	
Att_{ij}	Team j 's average attendance of season i
Independent variable	
$Population_{nij}$	Population of n^{th} county
$Income_{nij}$	Median household income for n^{th} county
$Distance_{nij}$	Distance between stadium and population center of n^{th} county
TQ_{ij}	Team j 's winning percentage of the season i
$Payroll_{ij}$	Average payroll of home team j
$FinalRank_{ij}$	Team j 's final rank of the season i
$Playoff_{ij}$	Dummy of home team playoff appearance
$StarPlayer_{ij}$	Team j 's number of star players
$TeamAge_{ij}$	Team j 's age
$Champs_{ij}$	Team j 's number of previous championships
STD_Age_{ij}	Stadium age of home team j
$Capacity_{ij}$	Stadium capacity of home team j
$Ticket_{ij}$	Average ticket price of the season i
$ProTeams_{ij}$	Number of other professional teams in same area
$Season_{ij}$	Number of seasons played from the 2006 season (Longitudinal term)
$Season_{ij}^2$	Square of $Season_{ij}$ variable (Quadratic term)

i : season i ($i = 1, \dots, 12$), j : team j ($j = 1, \dots, 29$), and n : county of n^{th} nearest from stadium ($n = 1, \dots, 5$)

minants of four categories (RQ3), the full model was constructed by including several Level I determinant variables with the stadium location variable, decided in previous demand-weighted location models, in the null model. Based on the result of this full model, more clarity was provided regarding which determinants had an effect on MLB attendance during the 12 seasons examined. In addition, the relationship of variances between seasons and between teams was examined. In addition, the final model with location variable was suggested by the -2 log likelihood ratio test of the full location model. All of these statistical models were analyzed by using SAS 9.4.

Results

Before examining the influence of MLB stadium location on attendance, the two-level null model was conducted to analyze the data structure of MLB attendance. Per the null model results, the variances of Level I (σ^{*2}) and II (τ_0^{*2}) were 17,422,451 and

49,450,062. Based on these values, the intraclass correlation (ICC) of season was calculated at 26.1%, indicating that nearly 26% of the total variance in MLB teams' average attendance could be attributed to the seasonal differences of each team. Also, the remainder (73.9%) was from the differences between teams. This variance structure of data shows the average attendance of MLB depended on team characteristic more than seasonal changes.

In addition, the variables (i.e., population, income, and population center) were used to examine the relationship between stadium location and MLB attendance, and the descriptive statistics of seasonal average attendance and neighboring counties near stadiums from 2006 to 2017 are illustrated in Table 2. The distance between the stadium and the county's population center determines the order of counties. The population of the nearest county is higher than that of other counties, while the income level of the nearest county is the lowest. In addition, the variation of distance increases slightly by the order of county because the

Table 2. Statistics of seasonal average attendance and neighboring counties

Variable	N	Mean	Std. Dev.	Min	Max	
Ave. ATT (per season)	348	30,873.0	8,189.3	14,372	53,069	
1st county	Distance (Km)	348	7.98	4.88	1.03	19.96
	Population	348	2,091,401	2,027,976	86,858	10,170,292
	Income	348	54,336	12,758	30,936	11,0816
2nd county	Distance (Km)	348	27.95	22.26	6.66	121.81
	Population	348	1,270,489	1,802,492	153,956	10,170,292
	Income	348	64,818	15,554	36,660	117,237
3rd county	Distance (Km)	348	38.31	28.81	10.04	131.66
	Population	348	818,141	722,703	51,441	3,190,400
	Income	348	65,251	16,669	31,494	120,941
4th county	Distance (Km)	348	47.35	35.95	14.84	152.42
	Population	348	627,449	639,629	37,895	2,648,771
	Income	348	65,229	16,826	30,602	118,279
5th county	Distance (Km)	348	57.98	44.09	14.92	181.58
	Population	348	808,512	1,791,592	31,752	10,170,292
	Income	348	66,502	13,088	42,984	113,908

county sizes by state vary. For example, the county areas of Arizona and southern California are relatively larger than those of other states. Also, the maximum number of collected counties' data is five because some counties overlap even though the straight-line distance is over 150 kilometers. For example, the fifth county of the San Diego Padres is Los Angeles, which is the home county of the Los Angeles Dodgers, and the county's distance is 180 kilometers. Also, the fifth county of Arizona, Pima, is 171 kilometers from Chase Field, the home of the Arizona Diamondbacks. Also, the values of population and income were down-scaled by ten to the negative third power (10^{-3}) because of high standard deviation by multiplying two variables.

These data were used for the analysis of the relationship between attendance and stadium location. This analysis was constructed by changing the power of distance variable and the number of counties. Also, the slope of Figure 1 and Figure 2 are the graphic results of the model fit comparison.

In the case of the model fit comparison by the power (x) of distance (Figure 1), x was changed from 0.1 to 10. The model fit statistics, such as -2 log likelihood, AIC, and BIC, indicate that the model fits become better by increasing the power of distance. However, the model fit results were converged, and their difference was gradually reduced after the square of distance ($x = 2$). Also, the location variable was a significant predictor of average attendance from the origin-

al value of distance ($x = 1.0$; $\alpha_1 = .0232$, $t = 2.03$, $p < .05$) to the 10th power ($x = 10$; $\alpha_1 = .0309$, $t = 2.08$, $p < .05$). Based on these results of the model fit comparisons and simplification of the equation, the optimal power (x) of distance was determined as two ($x = 2$) for the analysis of the distance effect between the stadium and the city on MLB attendance, and the location model of an MLB stadium was determined as follows:

$$Att_{ij} = \alpha_0 + \alpha_1 \left(\frac{P_{1ij} * I_{1ij}}{D_{1ij}^2} \right) + \epsilon_{ij}$$

Based on the results of the model above, the other model fit comparison was conducted for finding the optimal number (n) of counties to influence MLB attendance. Figure 2 shows that the model fits (i.e., -2 Log Likelihood when $x = 1, 2$, and 3) are stable when the number of counties is increased. Therefore, the optimal number of counties for the location variable was one ($x = 2$, $n = 1$; $\beta_1 = .0247$, $t = 2.11$, $p < .05$). This result showed the MLB attendance was principally influenced by the demographic and geographic factors of the county that stadium is located.

From the decision of the optimal number of counties, the demand-weighted location model of a stadium was determined as follows:

$$Att_{ij} = \beta_0 + \beta_1 Location_{ij} + \epsilon_{ij}$$

$$= \beta_0 + \beta_1 \left(\frac{P_{1ij} * I_{1ij}}{D_{1ij}^2} \right) + \epsilon_{ij}$$

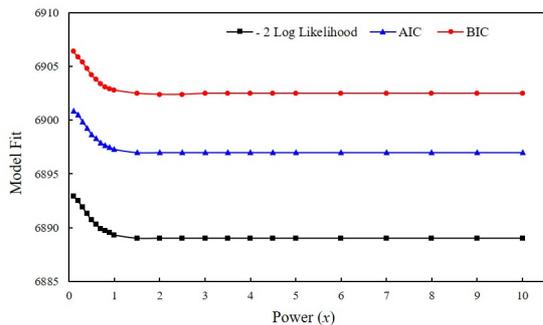


Figure 1. The model fit comparisons by the power (x) of distance

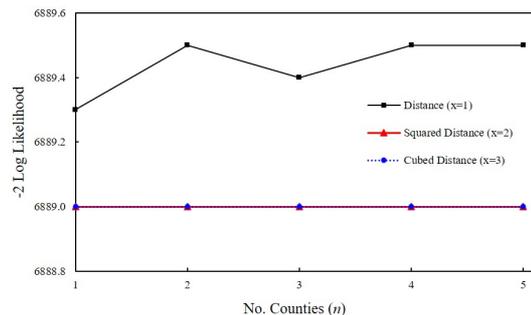


Figure 2. The model fit comparisons by the number (n) of counties

Based on the results of the optional location model above, 14.8% of the variance by differences between teams was reduced by location variables ($\tau_0^2 = 42,122,476$) while the variance by seasonal differences wasn't changed ($\sigma^2 = 17,445,840$) because most MLB teams had not changed the location of their home stadium during 2006-17 seasons. Even though six MLB teams (i.e., Atlanta Braves, Miami Marlins, Minnesota Twins, New York Yankees, New York Mets, and Washington Nationals) relocated to another stadium during this period, only the Braves and the Marlins moved to a new stadium in the different county.

In addition to this location variable, other attendance determinants in four categories were adopted in this study, and the descriptive statistics of these variables during the 2006 to 2017 MLB regular seasons are shown in Table 3.

Before the analysis of the relationship between MLB attendance and attendance determinants, including the location variable, the data centering by grand means was applied. Table 4 indicates the results of the null model, location model, full model, and final model.

From the likelihood ratio tests for fixed effects of the full models, the final models were suggested.

Among the 14 independent variables, 8 variables (i.e., *Location*, *FinalRank*, *Playoff*, *TeamAge*, *Champs*, *STD_Age*, *ProTeams*, and *Season*²) were excluded at the full model of MLB attendance model. Per these results, the final model of MLB attendance model was suggested as follows:

$$\begin{aligned} Att_{ij} = & \gamma_{00} + \gamma_{10}(TQ_{ij}) + \gamma_{20}(Payroll_{ij}) \\ & + \gamma_{30}(StarPlayer_{ij}) + \gamma_{40}(Capacity_{ij}) \\ & + \gamma_{50}(Ticket_{ij}) + \gamma_{60}(Season_{ij}) \\ & + R_{ij} + U_{0j} \end{aligned}$$

where, σ^2 : between seasons variance (R_{ij})

τ_0^2 : between teams variance (U_{0j})

In addition, Table 4 shows that the six independent variables (i.e., home team's final winning percentage, average payroll, and number of star players, stadium capacity, average ticket price, and number of seasons played) significantly influenced each team's seasonal average attendance from the 2006 to 2017 MLB regular seasons.

Especially, the season final winning percentage ($\gamma_{10} = 20210$, $t = 7.07$, $p < .001$), payroll ($\gamma_{20} = .0023$, $t = 11.87$, $p < .001$), and number of star player ($\gamma_{30} =$

Table 3. Statistics of 2006-17 MLB average attendance and attendance determinant variables

Variable	N	Mean	Std. Dev.	Min	Max
<i>Att</i>	348	30,873.0	8,189.3	14,372	53,069
<i>Location</i>	348	21,827.2	97,257.2	48.7	622,762.7
<i>TQ</i>	348	0.500	0.067	0.315	0.642
<i>Payroll</i>	348	3,617,192	1,504,755	576,865	8,253,336
<i>FinalRank</i>	348	3.0	1.5	1	6
<i>Playoff</i>	348	0.30	0.46	0	1
<i>StarPlayer</i>	348	2.0	1.4	0	7
<i>TeamAge</i>	348	82.2	43.4	8	141
<i>Champs</i>	348	3.6	5.2	0	27
<i>STD_Age</i>	348	22.7	24.8	0	105
<i>Capacity</i>	348	43,475	5,280	31,042	57,333
<i>Ticket</i>	348	27.17	9.68	13.71	72.97
<i>ProTeams</i>	348	2.7	1.4	1	6

Table 4. Results of null model, location model, full model, and final model

	Null Model (ML)		Location Model (ML)		Full Model (ML)		Final Model (ML)	
	Estimate	(SE)	Estimate	(SE)	Estimate	(SE)	Estimate	(SE)
Fixed Effect								
<i>Intercept</i>	30,873	(1324.9)	30,873	(1225.8)	33,924	(807.0)	34,019	(763.26)
<i>Location</i>			0.025*	(0.011)	- 0.001	(0.007)		
<i>TQ</i>					16,860**	(4,105)	20,210**	(2,858)
<i>Payroll</i>					0.0023**	(0.0002)	0.0023**	(0.0002)
<i>FinalRank</i>					205.2	(183.1)		
<i>Playoff</i>					-394.9	(139.1)		
<i>StarPlayer</i>					276.1*	(139.2)	284.4*	(138.7)
<i>TeamAge</i>					25.83	(18.54)		
<i>Champs</i>					-66.07	(159.24)		
<i>STD_Age</i>					-9.98	(20.10)		
<i>Capacity</i>					0.56**	(0.08)	0.55**	(0.07)
<i>Ticket</i>					70.00	(40.55)	79.68*	(39.44)
<i>ProTeams</i>					472.3	(544.9)		
<i>Season</i>					-434.6*	(167.1)	-570.8**	(57.7)
<i>Season²</i>					-12.7	(14.3)		
Random Effect								
Residual (σ^2)	17,422,451	(1,379,523)	17,445,840	(1,381,649)	7,762,172	(617,472)	7,871,308	(624,277)
τ_0^2	49,450,062	(13,367,997)	42,122,476	(11,469,213)	12,858,002	(3,725,540)	13,314,489	(3,735,297)
R-squared								
R_1^2			0.109		0.692		0.683	
R_2^2			0.144		0.735		0.726	
-2 Log Likelihood	6893.0		6889.0		6596.6		6602.1	

Standard errors are in parentheses.

* $p < .05$ and ** $p < .001$

284.4, $t = 2.05$, $p < .05$) of teams positively affected the seasonal MLB attendance. Furthermore, the effects of stadium capacity ($\gamma_{40} = 0.55$, $t = 7.57$, $p < .001$) and ticket price ($\gamma_{50} = 79.68$, $t = 2.02$, $p < .05$) on attendance were significant. However, the recent season's attendance had been gradually decreased by years ($\gamma_{60} = -570.8$, $t = -9.89$, $p < .001$).

To check the global fit of the final model, R^2 measures were adopted in this study. Based on the variance between seasons, and between teams, R_1^2 and R_2^2 were measured, and R_1^2 indicates that the propor-

tional reduction of prediction error by the final model six independent variables is 68.3%. Also, R_2^2 shows that the prediction error of MLB attendance would be reduced to 72.6% if season i is fixed and team j is randomly chosen, while R_1^2 and R_2^2 of the location model are 10.9% and 14.4%.

Discussion

The purpose of this study was to investigate the effect of stadium location on attendance for the opti-

mal MLB stadium location modeling. For this analysis, two-level hierarchical linear modeling was adopted because the structure of data had the feature of the nested data (i.e., 12 seasons and 29 MLB teams). Through the analysis of the variance structure by the level of difference, this nested data structure was confirmed. In addition to the two-level modelling approach, this study is the first attempt to investigate the optimal location of stadium related to attendance, and the location is expressed by population, income, and distance between stadium and population center. Even though several studies have adopted the distance from the home team stadium to the nearest other stadium (Winfree et al., 2004), and between the home team and the visiting team (DeSchriver & Jensen, 2002; Lemke et al., 2010), these variables have been investigated as an indicator of other options that might be available for viewing other games and rivalry games rather than as factors of stadium location, which was the focus of the current study. Nelson (2002) analyzed the influence of stadium location in downtown, edge, and suburban areas, but that study was focused on economic gains or losses rather than on the attendance factor examined in the present study.

From the results of location modeling in this study, the seasonal average of MLB attendance was found to be in inverse proportion to the squared distance. Such a finding reveals that distance should be considered as a significant variable, and a short distance between stadium and population center leads to increased attendance. This also suggests the importance of investigating the county in which a sporting facility is located. Although the outcome of the optimal locating modeling is the first attempt of stadium location modeling, it is an approximation and has limitations. Because the units of population, income, and population center in the present study are related to the county instead of the city or the smallest administrative district, and the size of the county varies by state, the results of the data of these slightly larger units and different sizes are limited in terms of accounting for

the attendance at all MLB stadiums. In addition, although this study did not include the results of the analysis that included income level and population of city instead of the location factor, the results revealed that income was a significant predictor of MLB attendance, whereas population was not. Because of the location factor, which was a combination of significant (i.e., income) and insignificant (i.e., population) variables, it could be regarded as an insignificant predictor of attendance in the final model.

Further, the current study confirmed the general fact that when teams are looking to possibly move into another stadium, deciding on the new stadium location teams should take into account the population and income of the area. This finding indicates that further study is necessary because of the limitations of the location factors mentioned above. For example, studies using uniform data with no regional differences in units smaller than the county are expected to be a more effective and accurate analysis of the effects of location factors. Also, it is necessary to investigate the location of individual stadiums to reflect local characteristics such as spectator demographics, traffic conditions, and parking conditions in each area. Furthermore, because the effects of the location of stadiums could vary depending on sports such as daily based (e.g., baseball, basketball), weekly based (e.g., soccer, football), and other periodical sporting events (e.g., motorsport), research on various sports is needed. Thus, future stadium location modeling investigations looking into other sports can build up on this study by analyzing other sports and sport industry situations.

In addition to the location factor, several other attendance determinants of the four major categories have also influenced attendance during MLB regular seasons. In the attractiveness factors, the home team's quality (i.e., winning percentage) and popularity (i.e., average payroll, and star players) influenced attendance from the 2006 to 2017 MLB regular seasons. Several studies have also emphasized the significant influence of the home team's quality and popularity on

attendance (e.g., Carmichael et al., 1999; Hansen & Gauthier, 1989). The results of the current study confirm the importance of the home team's quality and popularity as they both relate to increased attendance. However, some variables related to the home team's attractiveness were found to be insignificant predictors if those variables are minutely examined for a specific season. For instance, the home team's playoff appearance was the only significant predictor of attendance in the 2014 MLB season (Lim & Pedersen, 2018), and other home team attractiveness factors were not. Even though these variables were found to not have statistically influenced attendance on certain occasions, the findings of this study reveal that the variables related to the quality and popularity of the home team are significant predictors of the MLB attendance.

In the residual preference factors, the stadium capacity is a significant predictor of MLB attendance. Stadium capacity is regarded as an indicator of viewing quality (e.g., crowdedness). While some studies are unable to demonstrate the significant influence of stadium capacity on attendance with regard to the MLB (e.g., McDonald & Rascher, 2000), Ferreira and Bravo (2007) found that stadium capacity positively influenced Chilean soccer league attendance. With regard to stadium age, however, the current study was unable to explain the results of previous studies (e.g., Coates & Humphreys, 2007) that noted that the novelty effect of stadium encourages sports fans to attend sporting events. If the variable of stadium age in this study were reflected in the historical worth of stadiums such as Fenway Park (Boston Red Sox) and Wrigley Field (Chicago Cubs), the results would confirm the importance of novelty effect and historical meaning (Coates & Humphreys, 2007).

The elasticity of ticket price of sporting events has been a controversial issue for some time. The negative elasticity of ticket price in various sports leagues has been shown in several studies (e.g., Coates & Humphreys, 2007; Simmons, 1996), and these results accord closely with the microeconomic price theory

that is at the heart of the relationship between demand and supply. However, Baimbridge, Cameron, and Dawson (1995, 1996) noted different results with regard to English rugby and soccer. The current study also revealed the positive elasticity of ticket price in MLB attendance. To understand this positive elasticity, it is necessary to compare the attendance demands of 30 MLB teams because it is possible that the popular teams, such as the Red Sox, Yankees, and Cubs, have increased their ticket prices more so than other teams because their fans' high attendance demands warrant the price increase. The team popularity ranking (The Harris Polls, 2015) and the seasonal average attendance data from the 2006 to the 2015 MLB seasons suggest that the correlation of these two variables is .574. This correlation value would be sufficient to support the notion that high popularity leads to increased attendance, and this would in turn encourage popular teams to raise their ticket prices. The present study is limited with regard to its analysis of the relationship that might exist among demand, attendance, and ticket price because this is not the main purpose of the current study. In order to examine this relationship, additional data and a different research methodology would be necessary.

The season was adopted as variables of longitudinal terms and attendance in the years near the end of the study's timeframe gradually decreased. Because many MLB teams still depend on the revenue brought in through gate receipts, this decrease in attendance could expose MLB teams to financial risk. Therefore, the league as a whole (MLB) and the individual teams within the league should be concerned with this decrease in attendance, and a study whose purpose is to investigate this trend of decreasing attendance should be conducted as soon as possible.

This study adopted the 14 independent variables, including the longitudinal and quadratic variables (i.e., *Season*, and *Season*²) and the location variable, formulated by population, median income and distance, for an analysis of MLB attendance. However, some limi-

tations remain. As previously mentioned, the unit of location data should be small and consistent because there was a variation in the size of the counties examined in this study. Furthermore, the location modeling in this study did not reflect the specific accessibility conditions of each team (e.g., public transportations, parking availability and price, traffic and road conditions). Because these conditions vary with cities, investigating the stadium location of each team or city is something that would be of value in future studies. Moreover, the detailed and precise data related to population, income level and population center would contribute to an even more robust description of the relationship between attendance and stadium location.

Acknowledgments

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

References

- Aghezzaf, E. (2005). Capacity planning and warehouse location in supply chains with uncertain demands. *Journal of the Operational Research Society*, **56**(4), 453-462.
- Baade, R. A. (1996). Stadium subsidies make little economic sense for cities, a rejoinder. *Journal of Urban Affairs*, **18**(1), 33-37.
- Baimbridge, M., Cameron, S., & Dawson, P. (1995). Satellite broadcasting and match attendance: The case of Rugby League. *Applied Economics Letters*, **2**(10), 343-346.
- Baimbridge, M., Cameron, S., & Dawson, P. (1996). Satellite television and the demand for football: A whole new ball game? *Scottish Journal of Political Economy*, **43**(3), 317-333.
- Borland, J., & Macdonald, R. (2003). Demand for sport. *Oxford Review of Economic Policy*, **19**(4), 478-502.
- Campbell, J. R., & Hopenhayn, H. A. (2005). Market size matters. *The Journal of Industrial Economics*, **53**(1), 1-25.
- Carmichael, F., Millington, J., & Simmons, R. (1999). Elasticity of demand for Rugby League attendance and the impact of BskyB. *Applied Economics Letters*, **6**(12), 797-800.
- Cheng, E. W., Li, H., & Yu, L. (2005). The analytic network process (ANP) approach to location selection: A shopping mall illustration. *Construction Innovation*, **5**(2), 83-97.
- Coates, D., & Humphreys, B. R. (2007). Ticket prices, concessions and attendance at professional sporting events. *International Journal of Sport Finance*, **2**(3), 161-170.
- Current, J., Daskin, M., & Schilling, D. (2004). Discrete network location models. In Z. Drezner & H. W Hamacher (Eds.), *Facility location applications and theory* (2nd ed., pp. 81-118). New York: Springer.
- Current, J. R., & Schilling, D. A. (1987). Elimination of source A and B errors in p-median location problems. *Geographical Analysis*, **19**(2), 95-110.
- DeSchraver, T. D., & Jensen, P. E. (2002). Determinants of spectator attendance at NCAA Division II football contests. *Journal of Sport Management*, **16**, 311-330.
- Ferreira, M., & Bravo, G. (2007). A multilevel model analysis of professional soccer attendance in Chile 1990-2002. *International Journal of Sports Marketing and Sponsorship*, **8**(3), 49-66.
- Formby, B. (2016, October 29). In Arlington, voters debate paying to replace Rangers ballpark. *The Texas Tribune*. Retrieved from <https://www.texastribune.org/2016/10/29/rangers-ballpark-vote-arlington-hinges-faith-econo/>
- Francis, R. L., Lowe, T. J., & Tamir, A. (2004). Demand point aggregation analysis for a class of constrained location models: A penalty function approach. *IIE Transactions*, **36**(7), 601-609.
- Gratton, C., Shibli, S., & Coleman, R. (2005). Sport and

- economic regeneration in cities. *Urban Studies*, **42**(5), 985-999.
- Hakimi, S. L. (1964). Optimum locations of switching centers and the absolute centers and medians of a graph. *Operations Research*, **12**(3), 450-459.
- Hakimi, S. L. (1965). Optimum distribution of switching centers in a communication network and some related graph theoretic problems. *Operations Research*, **13**(3), 462-475.
- Hamacher, H. W., & Nickel, S. (1998). Classification of location models. *Location Science*, **6**(1-4), 229-242.
- Hansen, H., & Gauthier, R. (1989). Factors affecting attendance at professional sporting events. *Journal of Sport Management*, **3**, 15-32.
- Howard, D. R., & Crompton, J. L. (2014). *Financing sport* (3rd ed.). Morgantown, WV: Fitness Information Technology.
- Jeffe, J. (2013, November 11). Braves will move to Cobb County in shocking move. *Sport Illustrated*. Retrieved from <https://www.si.com/mlb/strike-zone/2013/11/11/atlanta-braves-moving-ballpark-cobb-county>
- Kemper, H. C., De Vente, W., Van Mechelen, W., & Twisk, J. W. R. (2001). Adolescent motor skill and performance: Is physical activity in adolescence related to adult physical fitness? *American Journal of Human Biology*, **13**(2), 180-189.
- Kimmelman, M. (2012, April 27). A ballpark that may be louder than the fans Marlins Park in Miami, baseball's newest stadium. *The New York Times*. Retrieved from <http://www.nytimes.com/2012/04/28/arts/design/marlins-park-in-miami-baseballs-newest-stadium.html>
- Leamer, E. E. (1968). Locational equilibria. *Journal of Regional Science*, **8**(2), 229-242.
- Lemke, R. J., Leonard, M., & Tlhokwane, K. (2010). Estimating attendance at Major League Baseball games for the 2007 season. *Journal of Sports Economics*, **11**, 316-348.
- Lim, N., & Pedersen, P. M. (2018). Examining determinants of sport event attendance: A multilevel analysis of a Major League Baseball season. *Journal of Global Sport Management*. doi: 10.1080/24704067.2018.1537675.
- McDonald, M., & Rascher, D. A. (2000). Does bat day make cents? The effect of promotions on the demand for Major League Baseball. *Journal of Sport Management*, **14**, 8-27.
- McEvoy, C. D., Nagel, M. S., DeSchrive, T. D., & Brown, M. T. (2005). Facility age and attendance in Major League Baseball. *Sport Management Review*, **8**(1), 19-41.
- MLB attendance report. (2017). *ESPN*. Retrieved from <http://espn.go.com/mlb/attendance>
- Nelson, A. C. (2002). Locating Major League stadiums where they can make a difference empirical analysis with implications for all major public venues. *Public Works Management & Policy*, **7**(2), 98-114.
- Perry, D. (2016, November 9). Election Day: Arlington voters approve new Texas Rangers ballpark worth \$1 billion. *CBS Sports*. Retrieved from <https://www.cbssports.com/mlb/news/election-day-arlington-voters-approve-new-texas-rangers-ballpark-worth-1-billion/>
- Perry, K. (2001). Professional sports attendance as a proxy for new stadium spillover benefits. *The Park Place Economist*, **9**, 62-70.
- ReVelle, C. S., & Eiselt, H. A. (2005). Location analysis: A synthesis and survey. *European Journal of Operational Research*, **165**(1), 1-19.
- ReVelle, C. S., Eiselt, H. A., & Daskin, M. S. (2008). A bibliography for some fundamental problem categories in discrete location science. *European Journal of Operational Research*, **184**(3), 817-848.
- Rivers, D. H., & DeSchrive, T. D. (2002). Star players, payroll distribution, and Major League Baseball attendance. *Sport Marketing Quarterly*, **11**(3), 164-173.
- Schofield, J. A. (1983). Performance and attendance at professional team sports. *Journal of Sport Behavior*, **6**, 196-206.
- Serra, D., & Marianov, V. (1998). The p-median problem

- in a changing network: The case of Barcelona. *Location Science*, **6**, 383-394.
- Siegfried, J. J., & Eisenberg, J. D. (1980). The demand for minor league baseball. *Atlantic Economic Journal*, **8**(2), 59-69.
- Simmons, R. (1996). The demand for English league football: A club-level analysis. *Applied Economics*, **28**(2), 139-155.
- Snyder, L. V., & Daskin, M. S. (2005). Reliability models for facility location: The expected failure cost case. *Transportation Science*, **39**(3), 400-416.
- The Harris Polls. (2015, July 9). *Yankees, Red Sox and Cubs Top List of America's Favorite Baseball Teams*. Retrieved from <https://theharrispoll.com/among-those-who-follow-the-sport-the-new-york-yankees-continue-the-over-two-decades-long-winning-streak-theyve-been-on-since-2003-coming-in-once-again-as-americas-favorite/>
- Tu, C. C. (2005). How does a new sports stadium affect housing values? The case of FedEx field. *Land Economics*, **81**(3), 379-395.
- Tucker, T. (2013, November 12). Braves plan to build new stadium in Cobb. *The Atlanta Journal-Constitution*. Retrieved from <http://www.ajc.com/sports/baseball/braves-plan-build-new-stadium-cobb/g4N4VC7nuSPUX62DykwQ9K/#>
- U.S. Census Bureau. (2010). *Centers of population*. Retrieved from <https://www.census.gov/geo/reference/centersofpop.html>
- U.S. Census Bureau. (2018). *American FactFinder*. Retrieved from <http://factfinder.census.gov/faces/nav/jsf/pages/index.xhtml>
- Vanreusel, B., Renson, R., Beunen, G., Claessens, A. L., Lefevre, J., Lysens, R., & Eynde, B. V. (1997). A longitudinal study of youth sport participation and adherence to sport in adulthood. *International Review for the Sociology of Sport*, **32**, 373-387.
- Whitson, D., & Horne, J. (2006). Underestimated costs and overestimated benefits? Comparing the outcomes of sports mega-events in Canada and Japan. *The Sociological Review*, **54**(2), 71-89.
- Winfree, J. A., McCluskey, J. J., Mittelhammer, R. C., & Fort, R. (2004). Location and attendance in major league baseball. *Applied Economics*, **36**(19), 2117-2124.
- Zimbalist, A. (1998). The economics of stadiums, teams and cities. *Review of Policy Research*, **15**(1), 17-29.