South African Journal for Research in Sport, Physical Education and Recreation, 2021, 43(3): 17 - 30.
Suid-Afrikaanse Joernaal vir Navorsing in Sport, Liggaamlike Opvoedkunde en Ontspanning, 2021, 43(3): 17 - 30.
ISBN: 0379-9069

USING HIERARCHICAL LINEAR ANALYSIS TO EXAMINE ATTENDANCE DETERMINANTS IN MAJOR LEAGUE BASEBALL

Namhun LIM[1], Wanyong CHOI[2], Paul M. PEDERSEN[3]
[1] *Department of Business, Accounting and Sport Management, Elizabeth City State University, Elizabeth City, NC, United States of America*
[2] *Department of Leadership, School Counseling and Sport Management, University of North Florida, Jacksonville, FL, United States of America*
[3] *Department of Kinesiology, Indiana University, Bloomington, IN, United States of America*

ABSTRACT

Although ticket sales revenues have been a major source of income in professional sports, Major League Baseball (MLB) has shown a steady decline in attendance over the past decade. Thus, it is necessary to investigate which attendance determinants have significantly affected the recent declining number of MLB spectators. Considering characteristics of the attendance data for multiple seasons, teams and games, a three-level Hierarchical Linear Model (HLM) was used to investigate the relationship between attendance determinants and the number of spectators across MLB seasons. Among the 13 game-level and 12 season-level attendance determinants adopted in this study, 12 game-level (e.g.visiting team's quality, championships, rivalry game) and five season-level (e.g.home teams' payroll, stadium capacity, ticket price) variables significantly affected MLB game attendance ($p<0.05$). The results revealed additional game-level determinants as significant variables, verifying how important it is to apply appropriate analytical methods depending on the data structure.

Keywords: Attendance; Baseball; Major League Baseball (MLB); Hierarchical linear modeling (HLM).

INTRODUCTION

Among the revenue streams in professional sports, ticket sales had long been a steady main revenue source. However, according to the report of 'PricewaterhouseCoopers' (PwC, 2019), media rights revenues have accounted for the largest revenue resources in the North American sports market since 2017. Additionally, because the compound annual growth rate of ticket sales revenue (2.5%) is smaller than that of media rights (4.6%) and sponsorship (3.8%), the revenue gap between media rights and ticket sales has been growing gradually. Rather, sponsorship revenues may replace ticket sales revenues, taking the second spot in the near future. While the portion of revenue from ticket sales is declining, the PwC report shows that ticket sales ($19.2 billion) still accounted for more than 27% of total revenue ($71.1 billion) in 2018, which makes it indisputable that ticket sales still make up one of the principal revenue sources in professional sports.

Because ticket sales revenues have been a major source of revenue in professional sports for such a long time, there have been various studies (Simmons, 1996; Coates & Humphreys, 2007) on the linkage between attendance at the sporting events and ticket sales revenues. In particular, many studies have been developed to ascertain which attendance determinants, including star players (Rivers & DeSchriver, 2002), facility age (McEvoy *et al*., 2005) and game uncertainty (Lee & Fort, 2008), significantly affect attendance in MLB. While the global pandemic made 2020 an exception, typically attendance is a key factor in professional baseball because MLB teams play more games per regular season (81 home games per season) compared to other sports leagues and the league, behind the National Football League (NFL), has the second highest average attendance among the North American professional sports leagues. Although various studies (Rivers & DeSchriver, 2002; Lee & Fort, 2008; Lim & Pedersen, 2018) have been conducted on MLB attendance, the league has recently shown a steady decline since the 2012 regular season (Gough, 2019). Thus, it is necessary to investigate which attendance determinants have significantly affected the recent declining number of MLB spectators.

Along with the need for research on the recent decrease in MLB attendance, a new methodological approach is also necessary. Many studies on attendance at sporting events (Rivers & DeSchriver, 2002; Lemke *et al.,* 2010) have identified a causal relationship between attendance and attendance determinants using a General Linear Model (GLM). If the data do not have hierarchical or nested data characteristics, a GLM is an acceptable analytical method to examine non-hierarchical data.

Considering the characteristics of the attendance data for multiple seasons and teams, the data vary depending on the differences between teams. They also vary depending on the differences between seasons and individual games, even if the data are from the same team and the same season for each team. Thus, the variance in hierarchical data is caused by the differences in each level (individual game-level, season-level, and team-level). While a GLM is unable to analyse the nested data consideration of these hierarchical characteristics, a Hierarchical Linear Model (HLM) has the advantage of analysing nested data, such as how much of the data variance for each level can be explained by independent variables (Snijders & Bosker, 2012).

Among the previous studies on attendance in baseball, some studies use a two-level HLM to examine individual game attendance for specific seasons (Lim & Pedersen, 2018) and the average attendance of each team for multiple seasons (Lim *et al*., 2019a). Based on such research, it is necessary to apply the three-level HLM (game-level, season-level and team-level) to analyse the number of individual game spectators for each MLB team's multiple seasons.

LITERATURE REVIEW

Attendance determinants in sporting events

Several studies have been conducted on attendance at sporting events (Simmons, 1996; Coates & Humphreys, 2007) because ticket sales revenues have long been a main revenue source for both amateur (intercollegiate football and basketball) and professional sports. In particular, Schofield (1983) introduced four categories of attendance determinants, namely economic, demographic, game attractiveness and residual preference factors. Many scholars (Hansen & Gauthier, 1989; Lee *et al*., 2003) have developed their research based on the Schofield classification.

Firstly, economic factors mean that various regional economic indicators, such as income level, ticket price and the existence of substitutional products or services, affect the number of spectators at sporting events. A certain income level leads to actual consumption behaviours

(Ahuvia & Friedman, 1998) and influences ticket purchases as a personal preference for sport fans to spend their leisure time (Hansen & Gauthier, 1989). In general economics, a higher income level leads to more consumption and, on the other hand, higher prices discourage consumers from spending. This negative price elasticity has been shown in the NFL (Welki & Zlatoper, 1994) and English soccer league (Simmons, 1996), while a study on English rugby (Baimbridge *et al*., 1995) shows different results.

Because sports consumption is sometimes expended by irrational decisions, unlike other products (Chadwick, 2006), it could lead to the conflicting results from the previous studies. Based on micro-economic price theory, in addition to the negative impact of high prices on consumption, the presence of substitutional products or services negatively affects the demand for purchases. If there are multiple professional sports teams in the same area, such as the New York Mets and the New York Yankees, the local sports fans can choose either option. Although the regular season is different for each sport, sporting events in different sports categories could be considered as a substitute. These negative effects have also been identified in previous studies (Baade & Tiehen, 1990; Baimbridge *et al*., 1995).

In addition to economic factors, demographic factors such as population influence attendance at sporting events. Generally, most major professional teams are located in populated cities because the market size and population size are positively related. Because of the importance of population as an attendance determinant, many previous studies have included population as one of the predictors of attendance and have shown that population has had a significant influence on attendance in MLB (Lemke *et al*., 2010), minor league baseball (Siegfried & Eisenberg, 1980) and the National Basketball Association (NBA) (Coates & Humphreys, 2007). With population, the ethnic mix of a team's home city has been considered as one of the demographic factors (Siegfried & Eisenberg, 1980) because the popularity of each sport varies according to race and ethnicity.

While economic and demographic factors are attendance determinants as regional information, the level of interest in the game (attractiveness factors) and the convenience and preference of attending (residual preference factors) influence the number of spectators who attend (Hansen & Gauthier, 1989). In particular, attractiveness factors are associated with the contents of individual games, such as game uncertainty, home and visiting teams' win-loss records, the number of star players, payroll, past achievements, league standing, rivalry games, and special promotions (Welki & Zlatoper, 1994). While attractiveness factors show how the degree of interest in a game affects sports fans' decisions on whether to attend, residual preference factors are related to external elements of games, such as stadium comfort (stadium capacity and age), day and time of a game and the weather (Lim & Pedersen, 2018).

Hierarchical data in sport

Most sport organisations and professional leagues, as well as amateur sports, have hierarchical organisational structures. Outcome data at the lower level of an organisation or league would be affected by variables at the same lower level, as well as variables at the higher level (Snijders & Bosker, 2012). For example, players' annual salaries may be affected by performance and popularity among sports fans or the public, but they may also be affected by the team to which the players belong. This kind of hierarchical data is often found in sports, and researchers are interested in investigating how independent variables at each level affect low-level dependent variables. For instance, Hofmann (1997) presents three possible options for analysing hierarchal data: (1) disaggregating the macro-level data to the micro-level; (2) aggregating the micro-level data to the macro-level; and (3) using hierarchical linear modeling (HLM). Hofmann (1997) points out that the first approach is not satisfied by the independence of observational assumption (Bryk & Raudenbush, 1992).

The second approach could ignore the variance of dependent variables by the micro-level variables. While the first and second approaches have disadvantages that are unable to be analysed by considering the separation of each level, HLM has the advantage of examining the variance of data by each level and the relationship between hierarchical levels (Snijders & Bosker, 2012). Therefore, HLM is more appropriate for analysing hierarchical data.

When examining the structure of attendance data, attendance at individual games is first influenced by differences between teams (team-level) and between seasons for each team (season-level). For each team's season, there is a variance of single-game attendance due to the differences between individual games or within a season (game-level). The structure of attendance data for individual games in a particular season could be divided into two levels, game-level and team-level (Lim & Pedersen, 2018), and the structure of seasonal average attendance data for each team can be separated by season-level and team-level (Ferreira & Bravo, 2007; Lim *et al*., 2019a).

PURPOSE OF RESEARCH

The purpose of this study is to investigate the relationship between the attendance determinants found in the previous, studies such as ticket price, income and population of home city, team's winning rate and league standing, rivalry games and stadium's age and capacity (Welki & Zlatoper, 1994; Rivers & DeSchriver, 2002; Coates & Humphreys, 2007) and the number of spectators in the 2009-18 MLB regular seasons using a three-level HLM. Based on the prior literature on attendance demand for sporting events and HLM, the present study pursued answers to the following research questions:

- RQ1: What is the data structure of the number of spectators in individual games during the 10 MLB regular seasons (2009-18)?
- RQ2: What is the relationship between significant attendance determinants found in previous research (Coates & Humphreys, 2007) and attendance during the 2009-18 MLB regular seasons?
- RQ3: How much could the attendance determinants in the final HLM model explain the variance of attendance within a season, between seasons and between teams?

METHODOLOGY

To investigate the data structure of attendance and the relationship between attendance and attendance determinants in MLB, the single game attendance for 30 MLB teams during the 2009-18 regular seasons (Sports Reference LLC, 2020) was used as a dependent variable (N=24,298).

Game-level attendance predictors

In this study, game-level attendance predictors were defined as independent variables with different values for each single game. The home team quality (*HTQ*) and visiting team quality (*VTQ*) were representative game-level variables and were calculated by the winning rates of the previous season (*j-1*) and the current season (*j*) prior to game *i* (Tainsky & McEvoy, 2012).

$$TQ_{ijk} = \left[Win\%_{(j-1)k} \times \left(162 - Progress_{(i-1)jk}\right) + \left(Win\%_{ijk} \times Progress_{(i-1)jk}\right)\right]/162$$

Where,

i: i^{th} home game of team *k* in season *j* (i=1 …, 81)
j: season *j* (j = 1, …, 10)
k: team *k* (k= 1, …, 30)
$Win\%_{(j-1)k}$: team *k*'s winning percentage in the previous season of season *j*
$Win\%_{ijk}$: team *k*'s winning percentage prior to game *i* in season *j*
$Progress_{ijk}$: team *k*'s number of games that have been played in season *j*, including game *i*

In addition, game uncertainty (*Uncertainty*) was calculated using Bill James's baseball game uncertainty equation below (Tainsky & McEvoy, 2012).

$$Uncertainty_{ijk} = |0.5 - C_{ijk}|$$

where,

$$C_{ijk} = (HTQ_{ijk} - HTQ_{ijk}VTQ_{ijk})/[\left(HTQ_{ijk} + VTQ_{ijk}\right) - 2HTQ_{ijk}VTQ_{ijk}]$$

Based on this uncertainty equation, a large difference in team quality between home and visiting teams increases the value of game uncertainty, while a similar team quality between the two teams lowers its value.

The variables of game back of division leader for the home team (*HT_GB*) and the visiting team (*VT_GB*) were adopted as the game-level independent variables. The visiting team's average payroll (*VT_Payroll);*(Major League Baseball Team, 2020), team age (*VT_Age*), number of championships (*VT_Champs*), and number of All-Star players (*VT_StarPlayers*) as reported by the official site of MLB (2020) were employed in this study. In addition, whether it was the match between divisional or local rivalries (*Rival)* (Lemke *et al*., 2010) and whether the game was played between Friday and Sunday (*Weekend*) were game-level attendance determinants. The season's progress (*Progress*) was used as a longitudinal term, indicating how many games the team played in season *j*, including game *i*. In addition, a quadratic term of the season's progress was adopted in the present study in the form of a square of the progress ($Progress^2$).

Season-level attendance predictors

Season-level attendance predictors could be defined as independent variables with different values for each season, with no significant difference or no difference available in an individual game. Attendance determinants related to the visiting team were used as predictors in the game level, while the home team's average payroll (*HT_Payroll*), team age (*HT_Age*), number of championships (*HT_Champs*) and number of star players (*HT_StarPlayers*) in season *j* were adopted as the season-level independent variable. Also, in the data provided by the U.S. Census Bureau (2019) along with data from Statistic Canada (2019) and the City of Toronto (2021) in order to include the only team outside the United States (Toronto Blue Jays), the home city's population (*population*) and median household income (*income*) for each team were used in the present study. Because the Minnesota Twins and the Tampa Bay Rays are located in two neighbouring cities (Minneapolis-St. Paul and St. Petersburg-Tampa), the population and income for the two teams used the sum of population and the population-weighted average income of the two neighbouring cities.

Table 1. **VARIABLE DESCRIPTIONS OF MLB ATTENDANCE DETERMINANTS**

Symbol	Description of variables
Dependent variable	
Att_{ijk}	Team *k*'s attendance of game *i* in the season *j*
Attendance determinants (Game-level)	
HTQ_{ijk}	Home team's winning percentage prior to game *i*
VTQ_{ijk}	Visiting team's winning percentage prior to game *i*
$Uncertainty_{ijk}$	Bill James's baseball game uncertainty
HT_GB	Home team's game back of division leader prior to game *i*
VT_GB	Visiting team's game back of division leader prior to game *i*
$VT\_Payroll_{ijk}$	Average payroll of visiting team
VT_Age_{ijk}	Visiting team's age
VT_Champs_{ijk}	Visiting team's number of previous Championships
$VT\_StarPlayers_{ijk}$	Visiting team's number of star players
$Rival_{ijk}$	Dummy of local or divisional rivalry games
$Weekend_{ijk}$	Dummy of weekend games (from Friday to Sunday)
$Progress_{ijk}$	Number of games that have been played, including game *i*
$Progress_{ijk}^2$	Quadratic term of $Progress_{ijk}$
Attendance determinants (Season-level)	
$HT\_Payroll_{\cdot jk}$	Total payroll of home team *k* in the season *j*
$HT\_Age_{\cdot jk}$	Home team *k*'s age
$HT\_Champs_{\cdot jk}$	Home team *k*'s number of previous championships
$HT\_StarPlayers_{\cdot jk}$	Home team *k*'s number of star players
$STD\_Age_{\cdot jk}$	Stadium age of home team *k*
$Capacity_{\cdot jk}$	Stadium capacity of home team *k*
$Ticket_{\cdot jk}$	Average ticket price change rate compared to previous season
$ProTeams_{\cdot jk}$	Number of other professional teams in same area
$Income_{\cdot jk}$	Median household income for home city
$Population_{\cdot jk}$	Population of home city
$Season_{\cdot jk}$	Number of seasons that have been participated from the 2009 season
$Season_{\cdot jk}^2$	Quadratic term of $Season_{\cdot jk}$

i: home game *i* (*i*=1, …, 81) *j*=season *j* (*j*=1, …, 10) *k*: team *k* in MLB (*k*=1, …, 30)

Additionally, the rate of change (*Ticket*) in average ticket prices (Team Marketing Report, 2020) compared to previous seasons and the number of other professional sports teams in the home city (*ProTeams*) were employed as economic factors of attendance determinants (Table1).

As the factors of residual preference, the age (*STD_Age*) and capacity (*STD_Capacity*) of stadiums for each MLB team (MLB, 2020) were adopted in the current study. Lastly, *Season* and *Season*^{2} were adopted as the season's longitudinal and quadratic terms.

After the data were collected, the unit was rescaled by dividing it by a thousand for variables with large units (i.e. *VT_Payroll*, *HT_Payroll*, *Income*, and *Population*). Also, the group means centering for the game-level variables and the grand mean centering for the season-level variables were used, except for binary variables (*Rival* and *Weekend*). The data centering and rescaling of these attendance determinants help to facilitate the understanding of the HLM results.

Data analysis

After collecting the data, the three-level null model was applied to examine the data structure of single game attendance in MLB (*RQ1*). The results of this null model show how much variation exists in MLB attendance data for each level. To examine the relationship between attendance and attendance determinants (*RQ2* and *RQ3*), the full model was constructed by including 13 game-level and 12 season-level attendance determinants in the null model. Moreover, the results of this full model provide the statistical evidence of which attendance determinants have significantly affected single game attendance during the 10 MLB regular seasons. The final model was proposed from the full model using the -2 log Likelihood ration test. Lastly, the proposed final model satisfied the assumptions of the hierarchical linear model (the normal distribution of residuals by the game-level and season-level independent variables and the normal distribution of season-level residuals for random coefficients). Also, SAS 9.4 was used for all these statistical models.

Ethical considerations

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors. Because the data employed in this study are public data (data published and available to the public), no separate permission is required to use the data. Also, because the research did not involve human participants, the study was not subject to Institutional Review Board (IRB) review and approval.

RESULTS

The descriptive statistics of variables in the present study are detailed in Table 2 to follow. In the case of attendance data, 97 out of 24,298 total games were excluded because the number of spectators for the first game of a doubleheader was not reported, and some games were played without spectators due to civil unrest in Baltimore in 2015. MLB's average attendance from the 2009 to 2018 regular seasons was 30,231, a difference of more than 6% from 28,339 in 2019 (Blum, 2019).

Table 3, to follow, shows the results of the three-level HLM null model for the MLB attendance data. The residual (R_{ijk}) represents the random effect of individual games that indicated the variance of MLB attendance due to differences between individual games within each season. At the same time, U_{ojk} and V_{ook} are the random effects of season *j* and team *k*, each a variance between seasons within each team, and between teams. The results of these random effects provide information about the data structure and show that 39.7% of the variance in the attendance of each individual game during the 10 MLB regular seasons (intraclass correlation

[ICC_{Game}]) could be attributed to between individual games within each season of each team. In addition, the variance of MLB attendance due to differences between teams (i.e., ICC_{Season}) accounts for 45.5% of the total variance, while the remainder (14.8%) of the variance (i.e., ICC_{Team}) could be due to differences between seasons for each team. Moreover, the random effects' results of the three-level HLM null model confirm the hierarchical structure of MLB attendance data and the need for the HLM approach to analyse these data.

Table 2. **DESCRIPTIVE STATISTICS OF MAJOR LEAGUE BASEBALL**

Variable	N	M±SD	Minimum	Maximum
Dependent Variable				
Att	24,201	30,231±10166.3	2,429	57,099
Independent Variables (Game-level)				
HTQ	24,298	0.500±0.064	0.288	0.683
VTQ	24,298	0.500±0.064	0.294	0.68
Uncertainty	24,298	0.076±0.053	0	0.332
HT_GB	24,298	6.50±8.82	-21	61
VT_GB	24,298	6.47±8.81	-21	60.5
VT_Payroll (US$)	24,298	3,903,867±1,525,431	896,438	8,253,336
VT_Age	24,298	82.55±43.46	11	142
VT_Champs	24,298	3.62±5.16	0	27
VT_StarPlayers	24,298	2.09±1.43	0	7
Rival	24,298	0.12±0.33	0	1
Weekend	24,298	0.48±0.50	0	1
Progress	24,298	81.50±46.76	1	163
Independent Variables (Season-level)				
HT_Payroll (US$)	24,298	3,904,187±1,525,772	896,438	8,253,336
HT_Age	24,298	82.57±43.46	11	142
HT_Champs	24,298	3.62±5.16	0	27
HT_StarPlayers	24,298	2.09±1.43	0	7
STD_Age	24,298	23.75±24.31	0	106
Capacity	24,298	43,256±5,114	31,042	56,000
Ticket (%)	24,298	3.10±11.26	-32.68	76.26
ProTeams	24,298	2.70±1.39	1	6
Income (US$)	24,298	49,986±14,737	23,600	123,767
Population	24,298	1,538,284±2,049,141	296,109	8,581,378

Note: This table shows the descriptive statistics results before data centering.

Table 3. **RESULTS OF NULL MODEL**

Effect	Estimate (ML)	Std. Error	ICC
Fixed Effect			
Intercept	30,208	1,272.4	
Random Effect			
Residual (σ^2)	41,014,267	375,182	0.397
τ_0^2	15,329,235	1,363,179	0.148
φ_0^2	46,989,685	12,542,272	0.455
-2 Log Likelihood	494,043.5		

ML = Using the maximum likelihood estimation

The results of the full model with the 13 game-level and 12 season-level independent variables added to the null model above are in Table 4. The study found that among the 13 game-level variables, all variables except for the number of star players on the visiting team ($p>0.05$) were significant attendance determinants affecting the number of spectators at MLB individual games. On the other hand, five of the 12 season-level variables (i.e.*HT_Payroll*, *HT_StarPlayers*, *STD_Capacity* [$p<0.001$], *Ticket*, and *Seasons*^{2} [$p<0.05$]) were found to be significant predictors of attendance.

To suggest the final model, variables determined to be insignificant predictors of attendance in the full model results were investigated through likelihood ratio tests. These insignificant variables (i.e *VT_StarPlayers*, *HT_Age*, *HT_Champs*, *STD_Age*, *ProTeams*, *Income*, *Population*, and *Season*) were excluded from the final model as they were determined not to affect the final model, even if excluded. Therefore, the final model was proposed as follows:

Game-level:
$$Att_{ijk} = \beta_{0jk} + \beta_{1jk}(HTQ_{ijk}) + \beta_{2jk}(VTQ_{ijk}) + \beta_{3jk}(Uncertainty_{ijk}) + \beta_{4jk}(HT\_GB_{ijk}) + \beta_{5jk}(VT\_GB_{ijk}) + \beta_{6jk}(VT\_Payroll_{ijk}) + \beta_{7jk}(VT\_Age_{ijk}) + \beta_{8jk}(VT\_Champs_{ijk}) + \beta_{9jk}(Rival_{ijk}) + \beta_{10jk}(Weekend_{ijk}) + \beta_{11jk}(Progress_{ijk}) + \beta_{12jk}(Progress^2_{ijk}) + R_{ijk}$$

Season-level:
$$\beta_{0jk} = \delta_{00k} + \delta_{01k}(HT\_Payroll_{.jk}) + \delta_{02k}(HT\_StarPlayers_{.jk}) + \delta_{03k}(STD\_Capacity_{.jk}) + +\delta_{04k}(Ticket_{.jk}) + \delta_{05k}(Season^2_{.jk}) + U_{0jk}$$

$$\beta_{njk} = \delta_{n0k} \text{ (When } n= 1, \ldots, 12)$$

Team-level:
$$\delta_{00k} = \gamma_{000} + V_{00k}$$
$$\delta_{0mk} = \gamma_{0m0} \quad \text{(When } m=1, \ldots, 5)$$
$$\delta_{n0k} = \gamma_{n00} \quad \text{(When } n=1, \ldots, 12)$$

Where,

- i: i^{th} home game in season j of team k ($i = 1,\ldots,81$)
- j: season j ($j = 1,\ldots,10$)
- k: team k ($k = 1,\ldots,30$)
- σ^2: variance between individual games (R_{ijk})
- τ_0^2: variance between seasons (U_{0jk})
- φ_0^2: variance between teams (V_{00k})

Table 4. **RESULTS OF FULL MODEL AND FINAL MODEL**

	Full model (ML)		**Final Model (ML)**	
	Estimate	Std. Error	Estimate	Std. Error
Fixed Effect				
Intercept	25,338	881.5	25,467	807.9
Independent Variables (Game-level)				
HTQ	17,684**	1683	17,526**	1682
VTQ	2,721*	907	2,780*	871
Uncertainty	1,673.4*	731.9	1694.3*	729.4
HT_GB	-117.23**	8.51	-117.72**	8.51
VT_GB	-32.12**	6.38	-32.34**	6.30
VT_Payroll[a]	0.547**	0.03	0.548**	0.030
VT_Age	5.38**	0.94	5.38**	0.94
VT_Champs	115.12**	8.48	115.08**	8.48
VT_StarPlayers	6.96	30.40		
Rival	1,010.46**	110.84	1010.99**	110.81
Weekend	5648.92**	69.19	5648.99**	69.19
Progress	79.70**	3.04	79.75**	3.04
Progress^2	-0.38**	0.018	-0.38**	0.018
Independent Variables (Season-level)				
HT_Payroll[a]	1.790**	0.168	1.856**	0.164
HT_Age	25.78	21.40		
HT_Champs	104.53	194.53		
HT_StarPlayers	400.43**	114.06	395.96**	115.10
STD_Age	-39.96	28.52		
Capacity	0.370**	0.093	0.373**	0.091
Ticket (%)	42.07*	13.10	41.69*	13.19
ProTeams	932.44	822.58		
Income[a]	0.489	32.475		
Population[a]	-0.301	0.593		
Season	58.66	186.74		
Season^2	-48.31*	20.51	-43.65**	6.20
Random effect				
Residual (σ^2)	28,880,375	264,194	28,880,263	264,192
τ_0^2	5,662,309	527,900	5,794,813	532,355
φ_0^2	17,297,261	5,232,679	17,617,762	4,825,376
- 2 Log Likelihood	485,369.0		485,375.3	
R-Square				
R^2	0.498		0.494	

[a] Unit was rescaled by dividing by a thousand. * p<0.05 ** p<0.001

Table 4 includes the results of the final model. These results show that the home team's quality (γ_{10k}=17526; p<0.001), game back of division leader (γ_{40k}= -117.7; p<0.001), and average payroll (γ_{01k}=1.86; p<0.001) have more influence on the number of spectators in MLB than the visiting team's variables (*VTQ* [γ_{20k}=2780; p<0.05], *VT_GB* [γ_{50k}= -32.34; p<0.001], and *VT_Payroll* [γ_{60k}=0.55; p<0.001]). The study also found that weekend games and rival games could expect an additional spectator amount of about 5,649 (p<0.001) and 1,011 (p<0.001) compared to weekday games and non-rival games. The negative influence of the quadratic term of the season on attendance (γ_{05k}=-43.7; p<0.001) is a statistical confirmation of the recent sharp decline in MLB attendance. Lastly, the final model could explain approximately 49% of the variance in the attendance data for MLB individual games (R^2=0.493).

DISCUSSION

The present study applied the first three-level Hierarchical Linear Modeling (HLM) to analyse the relationship between attendance determinants and attendance in baseball following previous HLM studies of individual game attendance for one season (Lim & Pedersen, 2018) and average spectators for multiple seasons (Lim *et al*., 2019a) using a two-level HLM. In addition, the results of random effects in this study show how much variation in the number of spectators account for the differences in each level in the individual games of MLB, confirming that the data structure of MLB attendance is hierarchical or nested. Because of such a hierarchical data structure, the findings confirm that HLM is appropriate as an analysis method of MLB attendance data.

The current study found that 12 of the 13 game-level variables had a significant impact on MLB attendance, while only five of the 12 season-level variables were considered significant. There are fewer variables at the season level than those at the game level that have such a significant influence, because the study analysed individual game attendance. If the data were analysed without hierarchical differentiation, more seasonal-level variables would become significant variables.

Because general linear modeling aggregates the lower-level data into the higher-level (Bryk & Raudenbush, 1992; Snijders & Bosker, 2012), the effects of the higher-level variables could be more overestimated than those of the lower-level variables. Indeed, the results of the current study's analysis that examined the same data used in this study using a generalised linear mixed model formulation with maximum-likelihood estimation showed that 10 season-level variables, excluding two variables (*HT_Champs* and *Season* [p>0.05]), were significant variables affecting the number of spectators in MLB. This is an example of how important it is to apply appropriate analytical methods depending on the data structure.

Some previous studies (Baimbridge *et al*., 1995) have shown positive price elasticity; that higher ticket prices could expect more spectators. However, Lim and Pedersen (2018) pointed out that further research that examines the relationships among the demand of attending, ticket price and actual attendance is necessary, because there is a possibility that the high demand has already reflected the price increase and that this high demand leads to increased attendance, even if the price is raised. In addition to the price already reflecting the demand for attendance, the use of average ticket prices as an independent variable would be to analyse the impact of comparative prices on the number of spectators by the other 29 MLB teams that are not considered for local prices.

In this study, the rate of change in price compared to the previous season was adopted as a variable, instead of the average ticket price to examine the influence on attendance through the price change. The study showed that the price change rate also has a positive effect on MLB attendance. As a result of the study, it could be concluded that price increases prompt an increase in the number of spectators, but it could be interpreted that each team sufficiently reflects their ticket price as the demand to visit the stadium increases. The relationship between attendance and ticket prices is determined by the complex relationship between price, demand and actual attendance, as noted in the previous study (Lim & Pedersen, 2018). Thus, a more specific study of price elasticity in attendance would explain the complex relationship between attendance and ticket prices.

PRACTICAL APPLICATION AND RECOMMENDATION

The study shows that attendance determinants found in previous studies (Coates & Humphreys, 2007) could account for about 49% of the variance in MLB attendance data (R^2=0.493). This value of R^2 is substantially higher than 29% of the research (Lim *et al*., 2019b) that analysed the NBA attendance in the recent 13 seasons using similar attendance determinants. The findings show that the attendance determinants used in this study could be more predictable of MLB attendance than the NBA.

Moreover, the current study confirmed that the influence of the home team-related attendance determinants (i.e.winning rate, game behind and average payroll) on the number of spectators is more important than those of the visiting team. Also, because these variables of the home team affect the spectators of all home games in 81 games, it is important to consider how to improve the team performance of the home team first, rather than the detailed ticket sales strategy of individual games depending on which team visits.

Considering the changes in various game rules to shorten MLB's playing time to revive attendance, MLB seems to regard playing time as a key factor in the decline in attendance. Because the current study did not apply game duration as an independent variable, further studies will need to determine whether game playing time negatively impacts the number of spectators. In addition, this study examined variables that have significantly affected the overall attendance of 30 MLB teams, regardless of which team had many spectators, such as the Yankees or the Los Angeles Dodgers, or which have a small number of spectators.

As stated in the results, about 45.5% of the variance in individual game attendance is due to differences between teams. By dividing groups according to the average attendance level, the analysis of the attendance determinants that affect the number of spectators in each group could help the corresponding teams in each group improve their attendance. Lastly, this study analysed MLB attendance using a total of 25 attendance determinants (i.e.13 game-level and 12 season-level variables). However, these variables could explain only about 49% of the variance in MLB attendance. To predict MLB attendance more accurately, additional attendance determinates have to be applied in further studies, including playing time to account for 51% of the variance that could not be explained in this study.

CONCLUSIONS

The present study confirmed the hierarchical structure of MLB attendance data and the need to use a Hierarchical Linear Model (HLM) approach to analyse the hierarchical data. In addition, this study used a three-level HLM to investigate which attendance determinants have been affecting the MLB attendance, which has been on a steady decline since 2012. The findings of the statistical analysis provide an explanation regarding the current attendance situation in MLB. The study's results should assist practitioners, who are looking to improve or maintain their team's attendance. Moreover, the present study provides a theoretical background for further research by introducing a more effective methodology that takes into account the characteristics of the unique attendance data.

REFERENCES

AHUVIA, A.C. & FRIEDMAN, D.C. (1998). Income, consumption, and subjective well-being: Toward a composite macromarketing model. *Journal of Macromarketing*, 18(2): 153-168.

BAADE, R.A. & TIEHEN, L.J. (1990). An analysis of Major League Baseball attendance, 1969-1987. *Journal of Sport and Social Issues*, 14(1): 14-32.

BAIMBRIDGE, M.; CAMERON, S. & DAWSON, P. (1995). Satellite broadcasting and match attendance: The case of Rugby League. *Applied Economics Letters*, 2(10): 343-346.

BLUM, R. (2019). MLB average attendance down 1.7% in 2019, marking fourth straight year of decline. *The Baltimore Sun*, 3 October. Hyperlink: [https://www.baltimoresun.com/sports/mlb/bs-sp-mlb-attendance-decline-20191003-ewkrfrpaajfpzkkx7mzhicc5i4-story.html]. Retrieved on 3 October 2019.

BRYK, A.S. & RAUDENBUSH, S.W. (1992). *Hierarchical linear models*. Newbury Park, CA: Sage.

CHADWICK, S. (2006). Soccer marketing and the irrational consumption of sport. *International Journal of Sports Marketing & Sponsorship*, 7(3): 3-3, [doi: org/10.1108/IJSMS-07-03-2006-B002].

CITY OF TORONTO (2021). Toronto at a Glance. Hyperlink: [https://www.toronto.ca/city-government/data-research-maps/toronto-at-a-glance/]. Retrieved on 5 January 2021.

COATES, D. & HUMPHREYS, B.R. (2007). Ticket prices, concessions and attendance at professional sporting events. *International Journal of Sport Finance*, 2(3): 161-170.

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.

GOUGH, C. (2019). *MLB average per game attendance 2009-2019*. Statista, 23 October. Hyperlink: [https://www.statista.com/statistics/235634/average-attendance-per-game-in-the-mlb--regular-season/]. Retrieved on 23 October 2019.

HANSEN, H. & GAUTHIER, R. (1989). Factors affecting attendance at professional sporting events. *Journal of Sport Management*, 3(1): 15-32.

HOFMANN, D.A. (1997). An overview of the logic and rationale of hierarchical linear models. *Journal of Management*, 23(6): 723-744.

LEE, S.; RYDER, C. & SHIN, H. (2003). An investigation of environmental motivation factors among Minor League Baseball fans. *Sport Journal*, 6(3). Hyperlink: http://thesportjournal.org/article/an-investigation-of-environmental-motivation-factors-affecting-fans-of-minor-league-baseball/]. Retrieved on 22 February 2019.

LEE, Y.H. & FORT, R. (2008). Attendance and the uncertainty-of-outcome hypothesis in baseball. *Review of Industrial Organization*, 33(4): 281-295.

LEMKE, R.J.; LEONARD, M. & TLHOKWANE, K. (2010). Estimating attendance at Major League Baseball games for the 2007 season. *Journal of Sports Economics*, 11(3): 316-348.

LIM, N.; CHOI, W. & PEDERSEN, P.M. (2019a). Analyzing the relationship between stadium location and attendance: A location modeling of Major League Baseball (MLB). *International Journal of Applied Sports Sciences*, 31(1): 70-85.

LIM, N.; KIM, W. & DRANE, D. (2019b). Predictive analysis of the attendance determinants of the National Basketball Association: A hierarchical linear regression modeling. *Journal of Contemporary Athletics*, 13(4): 209-224.

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*, December 2018. Hyperlink: [https://doi.org/10.1080/24704067.2018.1537675]. Retrieved on 31 December 2018.

MCEVOY, C.D.; NAGEL, M.S.; DESCHRIVER, T.D. & BROWN, M.T. (2005). Facility age and attendance in Major League Baseball. *Sport Management Review*, 8(1): 19-41.

MLB (Major League Baseball) (2020). *MLB.com | The official site of Major League Baseball*. Hyperlink: [https://www.mlb.com/]. Retrieved on 2 February 2020.

MLB TEAM PAYROLL TRACKER (2020). *Sportrac*. Hyperlink: [https://www.spotrac.com/mlb/payroll/]. Retrieved on 2 February 2020.

PWC (2019). At the gate and beyond: Outlook for the sport market in North America through 2023, November. Hyperlink: [https://www.pwc.com/us/en/industries/tmt/assets/pwc-sports-outlook-2019.pdf]. Retrieved on 1 December 2019.

RIVERS, D.H. & DESCHRIVER, 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(4): 196-206.

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.

SNIJDERS, T. A. & BOSKER, R. J. (2012). *Multilevel analysis: An introduction to basic and advanced multilevel modeling* (2nd ed.). London, UK: Sage.

SPORTS REFERENCE LLC (2020). *Team game-by-game schedule*. Baseball-Reference.com - Major League Statistics and Information. Hyperlink: [https://www.baseball-reference.com/teams/ARI/2018-schedule-scores.shtml]. Retrieved on 2 March 2020.

STATISTIC CANADA (2019). *Census profile, 2016 census.* Hyperlink: [https://www12.statcan.gc.ca/census-recensement/2016/dp-pd/prof/details/page.cfm?B1=All&Code1=3520&Code2=35&Data=Count&Geo1=CD&Geo2=PR&Lang=E&SearchPR=01&SearchText=Toronto&SearchType=Begins&TABID=1]. Retrieved on 21 September 2019.

TAINSKY, S. & MCEVOY, C.D. (2012). Television broadcast demand in markets without local teams. *Journal of Sports Economics*, 13(3): 250-265.

TEAM MARKETING REPORT (2020). *MLB fan cost index*. Hyperlink: [https://teammarketing.com/fci/2018-mlb-fan-cost-index-details-3-2-increase/]. Retrieved on 23 April 2020.

U.S. CENSUS BUREAU (2019). *QuickFacts*. Hyperlink: [https://www.census.gov/quickfacts/fact/ table/US/]. Retrieved on 1 July 2019.

WELKI, A.M. & ZLATOPER, T.J. (1994). US professional football: The demand for game-day attendance in 1991. *Managerial and Decision Economics*, 15(5): 489-495.

Corresponding author: Dr. Wanyong Choi; **Email:** w.choi@unf.edu

(*Subject editor: Prof. Maya van Gent)*