



## Estimating the Direction and Magnitude of Substitution Effects in MLB Event Attendance

Kim YS<sup>1\*</sup>, Moffitt C<sup>2</sup>

DOI:10.31033/IJEMR/16.1.2026.1840

<sup>1\*</sup> Yong Seog Kim, Professor, Data Analytics and Information Systems Department, Utah State University, USA.

<sup>2</sup> Clay Moffitt, Data Analysis Consultant, New York City, USA.

We estimate the magnitude of the negative impacts (or substitution effects) caused by NHL or NBA games with schedule conflicts on the attendance demand of MLB games. In particular, we aggregate such substitution effects over three different factors such as temporal aggregation factors (e.g., years and days of the week), team aggregation factor (e.g., each MLB team), and team performance aggregation factor (e.g., each MLB team's performance). Overall, we observe that NHL and NBA games that have schedule conflicts with MLB games more significantly negatively impact the attendance on MLB games during the weekdays (in particular, Thursday and Wednesday) than MLB games during the weekend. We also observe that MLB teams in high standings suffer less from negative substitution effects than MLB teams in low standings in the division. In particular, when MLB teams are in a losing streak, spectators show their deepest disappointments in their teams and avoid to attend MLB games in stadium.

**Keywords:** Substitution Effect, Sports Economics, Louis-Schmeling Paradox, Temporal Substitution Effect, Team Loyalty Substitution Effect, Team Performance Substitution Effect

Corresponding Author	How to Cite this Article	To Browse
Yong Seog Kim, Professor, Data Analytics and Information Systems Department, Utah State University, USA. Email: <a href="mailto:yong.kim@usu.edu">yong.kim@usu.edu</a>	Kim YS, Moffitt C, Estimating the Direction and Magnitude of Substitution Effects in MLB Event Attendance. Int J Engg Mgmt Res. 2026;16(1):84-94. Available From <a href="https://ijemr.vandanapublications.com/index.php/j/article/view/1840">https://ijemr.vandanapublications.com/index.php/j/article/view/1840</a>	

<b>Manuscript Received</b> 2026-01-01	<b>Review Round 1</b> 2026-01-16	<b>Review Round 2</b>	<b>Review Round 3</b>	<b>Accepted</b> 2026-02-04
<b>Conflict of Interest</b> None	<b>Funding</b> Nil	<b>Ethical Approval</b> Yes	<b>Plagiarism X-checker</b> 3.95	<b>Note</b>
 © 2026 by Kim YS, Moffitt C and Published by Vandana Publications. This is an Open Access article licensed under a Creative Commons Attribution 4.0 International License <a href="https://creativecommons.org/licenses/by/4.0/">https://creativecommons.org/licenses/by/4.0/</a> unported [CC BY 4.0]. 				

## 1. Introduction

According to Forbes' estimate, the overall revenue of the professional Major League Baseball (MLB) in US was \$11.34 billion in 2023 (an average of \$378 million per team), which makes MLB franchises the third-most worth on average after the National Football League (NFL) and the National Basketball Association (NBA) (Forbes & Statista, 2024). Furthermore, the MLB continues to be one of the most profitable sports leagues in the world, with the league-wide revenue of MLB franchises has almost doubled over the past ten years.

Note that there are largely two main sources for the MLB franchises to generate revenue: one with fan-related sales and marketing (e.g., ticket sales and merchandise sales) and the other with corporate-related sales and marketing (e.g., broadcast rights and sponsorships). For example, the financial success of MLB franchises in US is significantly impacted by the league's TV viewership from more than 9 million fans on seasonal and championship games, contributing more than 10% of the total overall revenue of the MLB league in 2023 (Forbes & Statista, 2024).

Regardless of aforementioned revenue sources, however, it is the fans and spectators who are strongly interested in watching games of their favorite teams either in stadiums or on television, and thus ultimately determine the financial success of the MLB franchises. In particular, the MLB franchises in US is well known to have the largest number of seasonal spectators ( $\geq 70$  million) among all sports league in the world (Forbes & Statista, 2024). In short, the attendance demand on MLB games during the season is one of the most important indicators to measure the financial success of MLB teams and franchises.

Therefore, it has been a very critical task for industry decision makers (e.g., regulators, executives, and administrators) and researchers in sports economics and marketing community to calibrate models to either estimate the attendance demand or identify factors that affect the attendance of MLB games, which will consequently determine the revenue of teams (Borland & MacDonald, 2003; Martins & Cro, 2018). Once such factors are identified and visualized, executives and administrators can tune marketing campaigns to further promote the revenue of teams and league.

Note that numerous studies have already identified various factors such as internal (e.g., psychological factor of spectators), external (e.g., weather or geographical distance between the cities where the teams are located), and economic factors (the income per capita of the home team's and the visiting team's region). Unlike prior studies, we intend to estimate and visualize both the direction and the magnitude of substitution effect caused by competing sport events on the attendance of MLB games. This substitution effect on the attendance demand of MLB games is often explained in terms of opportunity cost caused by the fact that spectators can physically attend only a sport event among available sport games due to schedule conflicts.

To this end, we consider games of two other professional sports leagues, NBA and the National Hockey League (NHL), as substitutes to MLB games. NFL games are not considered mainly because NFL season schedule is not overlapped with MLB season schedule. In contrast, NBA and NHL games can be accidentally scheduled on the same day at near location when and where MLB games are scheduled. Under such circumstances, we speculate that the attendance demand of MLB games is somewhat impacted as the opportunity cost of attending to MLB games varies due to alternatively available NBA and NHL games.

In this study, we plan to measure the direction and magnitude of substitution effects across three dimensions. The first dimension we consider is a temporal dimension (e.g., the day of the week or play time of the game), which is most likely to affect the direction and substitution of substitution effects. The substitution effects on this temporal dimension will reflect the aggregated effects over the entire MLB franchises by the definition. On the second dimension, the substitution effects are aggregated over each MLB team based on team specific quality (e.g., fans' relative loyalty toward MLB games over NBA and NHL games). Main insights from this dimension will be garnered by comparing substitution effects between MLB teams with and without high loyalty toward MLB games. Finally, the last dimension to assess substitution effects is to measure substitution effects based on teams' current ranks in the league or performance in recent games (e.g., winning or losing streak).

## 2. Literature Review

Identifying factors and calibrating models to predict the attendance demand for various sport events have garnered great attention from regulators, executives, and administrators of professional sports leagues as well as academic researchers. To this end, few researchers have directly analyzed the effective factors on the attendance of MLB franchises (Lemke et al., 2010), while many researchers studied other sports events such as soccer games (Coats & Humphreys, 2012; Garcia & Rodriguez, 2002; Hart et al., 1975; Villa et al., 2011). One of the most classical hypotheses to explain the attendance on sport events is based on the seminal work of Rottenberg (1956), implying that a more balanced sporting competition attracts more spectators due to higher outcome uncertainty (Cox, 2018; Forrerst & Simmons, 2002; Knowles et al., 1992). In essence, this so-called Louis-Schmeling paradox is to emphasize the competitive balance among teams so that the appropriate level of differences in the quality of rivals can be maintained (Neale, 1964). For the same reason, two other studies (Reilly, 2015; Martins & Cro, 2018) also claimed that rivalry games attract more spectators. Several other studies emphasized the very nature of the team such as team quality (Garcia & Rodriguez, 2002) and team performance (Dubin, 2001) as contributing factors of the attendance demand.

In contrast, many other studies found that the psychological factor of spectators who strongly want to watch their home teams win is an effective determinant of the attendance demand. For example, Lemke et al. (2010) claimed that attendance to MLB games increases as the chance of the home team winning the game increases. Other researchers also reported that many spectators prefer to attend sport events when their teams play a much inferior team (Buraimo & Simmons, 2008; Pawlowski & Anders, 2012). Another group of researchers identified external factors as effective determinants of the attendance demand. For example, Serrano et al. (2015) reported that the quality of competing teams such as the market value of the players can have a positive impact on the attendance. In another study (Czarnitzki & Stadtmann, 2002), team-specific values like supporter clubs and reputation are emphasized, which may explain why team-specific

attendance prediction models predict better than a generalized model (Strnad et al., 2017). Other studies identified scheduling (Forrest & Simmons, 2006), geographical distance between the cities where the teams are located (Buraimo & Simmons, 2008), or the population of hometown and league membership period of a team (Dobson & Goddard, 2011) as effective determinants of the attendance.

In terms of methodologies, many studies have adopted various tools from classical linear regression models to latest artificial intelligence and machine learning methods in sports economics and sports analytics fields. For example, the classical types of artificial neural networks (ANNs) and decision trees (DTs) have been widely adopted for various model tasks in sport literature (Maszczyk et al., 2011; McCullagh, 2010; Sahin & Erol, 2017).

Other studies combined multiple models into a collective prediction model called as ensemble as their analytics tools. Note that popular ensemble models such as Bagging (Breiman, 1996), AdaBoost (Freund & Schapire, 1995) or Random Forest (Breiman, 2001) have been known to significantly outperform the performance of single machine learning model in terms of its accuracy by reducing variance and bias components of errors from single prediction model. For example, King et. al. (2018) compared a set of machine learning algorithms to predict NHL average home game attendance with game- and team-level data. Interestingly, they included the number of Twitter followers as a surrogate for team popularity and recent team performance (e.g., win-loss streak). Similarly, Pand and Wang (2024) examined the attendance of NFL games over 5000 regular season games using various algorithms and reported that ensemble models like Random Forest outperformed other models using stadium related variables (e.g., name and age) and personal income as key predictive variables.

The most relevant set of previous studies to this paper, however, focuses on the economic factors of the attendance demand. For example, several studies tried to estimate the demand function based on ticket pricing (Garcia & Rodriguez, 2002; Madalozzo & Villar, 2009), the purchasing power of spectators measured by the income per capita of the home team's and the visiting team's region (Garcia & Rodriguez, 2002; Pand & Wang, 2024), or market size (Buraimo et al., 2009). In a recent study (Park et al., 2024), a macroeconomic variable,

unemployment rates, was found to be the most influential factor in ensemble models such as Random Forest and XGBoost to predict MLB game attendance using daily MLB game data between 2014 and 2019.

One of very few existing studies on substitution effects across different sports is Garcia and Rodriguez (2002) who explicitly considered the opportunity cost of the consumer by including a dummy variable which provides information about whether or not the game is played on a weekend. In their study, it was assumed that the demand on the sport event on a weekday decreases because the opportunity cost of attending the event increases. In another study (Wallrafen et al., 2019), it was reported that geographical proximity and scheduling overlaps would cause significantly negative substitution effects between top and lower division within the same sport in Germany. In their follow-up study, Wallrafen et al. (2022) studied fan substitution effects across different sports (e.g., handball, basketball, ice hockey and football league) even for games that are played a few days before or after.

### 3. Conceptual Developments

The central theme of this study is to estimate the direction and magnitude of substitution effects on the attendance demand of MLB games by the opportunity cost caused by the fact that spectators have to attend only a sport event in the presence of other sport games (e.g., NBA and NHL games). In particular, we plan to measure substitution effects on the following three dimensions.

On the first dimension, we consider substitution effects caused by a temporal factor such as the day of the week, play time and year of the game. For example, when NBA and NHL games are scheduled on the same day at near location when and where MLB games are scheduled, we speculate that the attendance demand of MLB games decreases because the opportunity cost of attending to MLB games increases (Garcia & Rodriguez, 2002). In addition, the negative substitution effects by conflicted NBA and NHL games are speculated to be more prominent on weekdays than on weekends mainly because spectators have more free times during weekends than weekdays, making the opportunity cost of watching MLB games during weekdays is more expensive if all other things being equal.

Note, however, that the substitution effect considered in this study is different from the substitution effect in consumer choice theory, which relates a change in the relative price of a good to the amount of that good demanded by a consumer (Varian, 2014).

On the second dimension, the substitution effects are aggregated based on team's qualitative measure (e.g., fans' relative loyalty toward MLB games over NBA and NHL games). The direction and magnitude of substitution effects on this dimension is difficult to be speculated in advance because fans' relative loyalty is strongly dependent on state-, teams-specific, and other economic and demographic factors. Therefore, the analysis on this dimension will be mainly explorative.

Finally, the last dimension we consider to assess substitution effects includes teams' current ranks in the league or performance in recent games (e.g., winning or losing streak). We first speculate that MLB teams with lower standings will suffer from more severe negative substitutions effects than teams with higher standings. We attribute this reasoning to the fact that spectators are more likely to attend to sports events when the chance of their home team winning the game is high (Lemke et al., 2010; Pawlowski & Anders, 2015). In addition, we posit that MLB teams in a long winning or losing streak suffer from more severely than other MLB teams. This is mainly because spectators are more likely to attend to sports events when the game between teams with rivalry relationships or similar ranks in the division is expected to be fun and exciting (Buraimo & Simmons, 2008; Neale, 1964).

In addition, we speculate that the substitution effect is affected by measures that reflect teams' current ranks in the league or performance in recent games.

### 4. Data Sets and Data Engineering

We downloaded several data sets through a data scraping program created with Python from multiple sources. First, the attendance data sets were scrapped from [www.baseball-reference.com](http://www.baseball-reference.com) for each MLB game between 2008 and 2018. These data sets result in a total of 53,452 records with several variables such as game date, game day of the week, home and away team names, result, day/night game indicator, scores of home and away teams, and so on.

Next, we collected the schedule information of NHL and NBA games between 2008 and 2018 to indicate which MLB games had schedule conflicts with either NHL or NBA games. These data sets were scraped from [www.hockey-reference.com](http://www.hockey-reference.com) for NHL data (a total of 14,027 records with information of game date, visiting and home team names, goal scores, attendance, and event city name) and from [www.basketball-reference.com](http://www.basketball-reference.com) for NBA data (a total of 14,219 records with the same information as in NHL data). We summarize the initial data sets with the list of variables in Table 1.

**Table 1:** List of Variables

Variable	Description
Game ID	Unique identifier with date, (home) team, and attendance
Date	Date of the MLB game
Season	Year between 2008 and 2018
Day of the Week	Day of the week of the MLB game
Location City	City where the MLB game was played
(Home) Team	MLB team that hosted the MLB game
Home/Away Indicator	Indicator if the MLB game was a home/away game
Opponent	Away team name
Result	Outcome of the game: Win, Loss, or Tie
Runs For	Number of runs team scored
Runs Against	Number of runs opponent scored
Record	Record of team performance after game
Place In Division	Ranking of the team in the same division
Games Behind	Number of games that the team is behind the division leader in the standings
Duration	Length of the MLB game in hours
Night/Day	Indicator if the MLB game was played in the night or day time
Attendance	Reported number of spectators
Streak	Numerical representation of team's winning or losing streak
NBA Conflict	Indicator if NBA games have schedule conflicts with the MLB game
NHL Conflict	Indicator if NHL games have schedule conflicts with the MLB game
Total Conflict	Indicator if NBA or NHL games have schedule conflicts with the MLB game

Based on initial data sets, we engineer several variables for this study. We first create Boolean indicators to mark whether each MLB game has a schedule conflict with either NHL or MBA games. Out of 53,452 MLB games, we found that 1,294 games were scheduled on the same date with NHL or MBA games. Next, we compute the average attendance (denoted as  $Attd_{Avg}(t, wd)$ ) of MLB games for all the possible combinations of each MLB

team (denoted as  $t$ ) and each day of the week (denoted as  $wd$ ). Then we use  $Attd_{Avg}(t, wd)$  values to estimate the direction (i.e., positive or negative) and magnitude (i.e., large or small) of substitution effects due to conflicted NHL or NBA games bring. To estimate the substitution effect of either NHL or MBA game scheduled on the same day with a MLB game, we compute the attendance deviance for a chosen game  $g$  (denoted as  $Attd_{Dev}_g$ ) by subtracting  $Attd_{Avg}(t, wd)$  from the attendance of a MLB game (denoted as  $Attd_g$ ) as shown in Equation (1).

$$Attd_{Dev}_g = Attd_g - Attd_{Avg}(t, wd) \quad (1)$$

Note that a positive (a negative) value of  $Attd_{Dev}_g$  represents an increase (decrease) of attendance to a specific MLB game that has a schedule conflict. Then, we compute the proportion of attendance deviance out of the average attendance (denoted as  $P(Attd_{Dev}_g)$ ) as follows:

$$P(Attd_{Dev}_g) = Attd_{Dev}_g / Attd_{Avg}(t, wd) \quad (2)$$

While we may use Equation (2) to estimate substitution effect of a specific MLB game, we do not intend to estimate the substitution effect for each MLB game with schedule conflict because such an estimate is not reliable. Instead, we intend to estimate substitution effects over multiple MLB games by aggregating factors (denoted as  $F$ ) such as each day of the week, each year, or each team. So, the substitution effect (denoted as  $Sub$ ) over an aggregating factor  $F$  due to a schedule conflict league (denoted as  $L \in \{NHL, MBA, \text{ or } Both\}$ ) is computed as follows:

$$Sub(F)^L = || Attd_{Dev}^L_g \subseteq F ||^- / || D(F) ||^L \quad (3)$$

where  $|| Attd_{Dev}^L_g \subseteq F ||^-$  represents the total number of MLB games that have schedule conflicts with  $L$  league games, are a subset of MLB games aggregated over a factor  $F$ , and have a negative value of  $Attd_{Dev}_g$ . Similarly,  $|| D(F) ||^L$  represents the total number of MLB games that have schedule conflicts with  $L$  league games and are a subset of MLB games aggregated over a factor  $F$ . In essence,  $Sub(F)^L$  in Equation (3) simply presents the proportion of MLB games with decreased attendance when they have schedule conflicts with attendance, when they have schedule conflicts with NHL or NBA games. For example, the substitution effect of NHL

games that have schedule conflicts with MLB games in 2018 (i.e.,  $Sub(2018)^{NHL}$ ) is computed by taking the proportion of the total number of 2018 MLB games with decreased attendance due to schedule conflicts with NHL games (i.e.,  $|| Attd\_Dev^{NHL}_g \in 2018 ||^-$ ) out of the total number of 2018 MLB games that have schedule conflicts with NHL games (i.e.,  $|| D(2018) ||^{NHL}$ ).

### 5. Findings and Discussion

#### Substitution Effects with Temporal Aggregation Factors:

In this subsection, we present the substitution effects over temporal factors, day of the week and year of the game. To this end, we computed the average attendance of all MLB games between 2008 and 2018 for each day of the week. Our finding was consistent with the common belief that weekend games would attract more spectators than weekday games: On average, Saturday games recorded the largest attendance (34,763) followed by Sunday (32,129), Friday (32,008), Thursday (27,802), Monday (27,433), Wednesday (27,313) and Tuesday (26,993).

Then we used Equations (1) through (3) to compute a set of  $Sub(day\ of\ the\ week)^L$  values, the set of substitution effects caused by NHL, NBA, or any one of these league games with schedule conflicts ( $L \in \{NHL, MBA, or Any\}$ ) for each day of the week as an aggregating factor. Our findings in Figure 1.

According to Figure 1, the negative substitution effects by conflicted NHL games were more prominent on weekdays in order of Thursday (68%, indicating that 68% MLB games on Thursday with conflicted schedules with NHL games resulted in the decrease in attendance) followed by Wednesday (65%) and Friday (62%) than on weekends (Saturday (52%) and Sunday (57%)). The negative substitution effects by conflicted NBA games followed a similar pattern, more prominent on weekdays (Wednesday (69%), Tuesday and Thursday (65%)) than on weekends (Saturday (59%) and Sunday (51%)).

These make sense from the spectators' perspective of opportunity costs. For example, most spectators have more free times during weekends than weekdays, making the opportunity cost of watching MLB games during weekdays is more expensive if all other things being equal. Therefore, it is very likely that spectators prefer watching NHL or NBA games during weekdays to watching MLB games during weekdays, which makes negative substitution effects during weekdays from NHL or NBA games on the attendance demand on MLB games much greater than those during weekends. We also find that Friday suffered less from the negative substitution effect by either conflicted NHL or NBA games than other weekdays. We attribute this finding to the fact that the opportunity cost of Friday is lower than those of other weekdays because it is the beginning of the weekends.

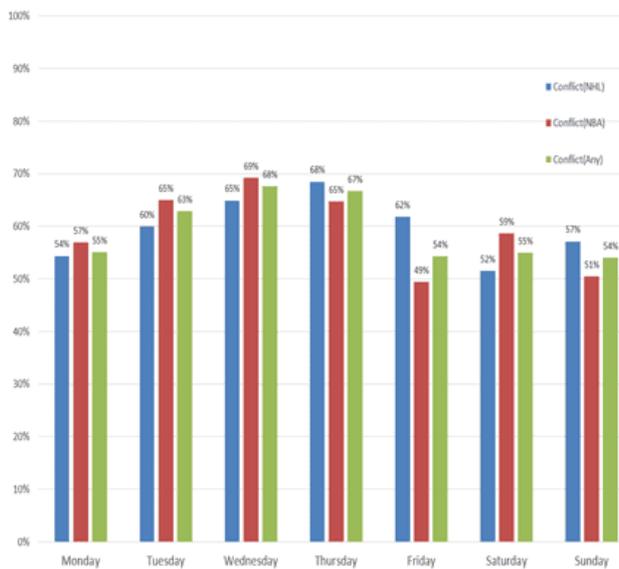


Figure 1: Substitution Effect of NHL and NBA Leagues over Days of the Week

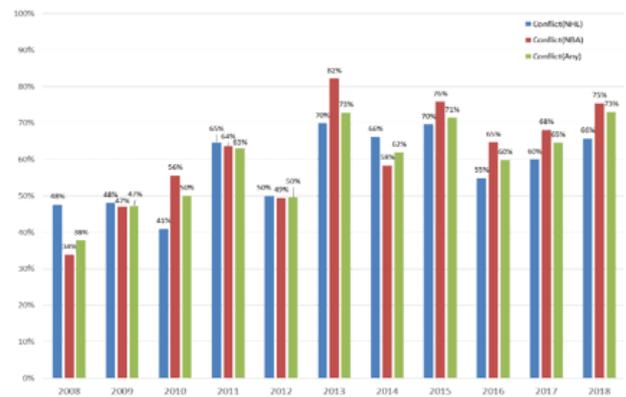


Figure 2: Substitution Effect of NHL and NBA Leagues over Years

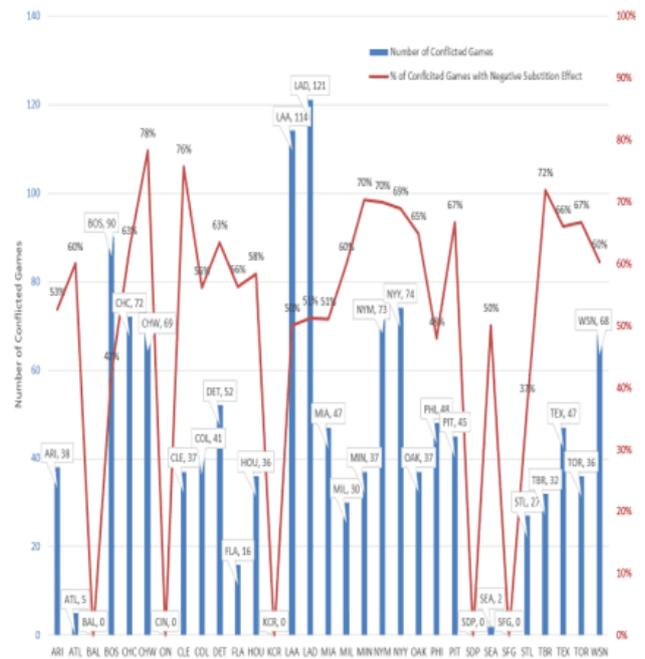
Similarly, we aggregate the substitution effects of conflicted NHL or NBA games over another temporal aggregation factor, year, and summarize them in Figure 2.

According to Figure 2, between 41% (2010) and 71% (2013 & 2015) of MLB games that had schedule conflicts with NHL games experienced a decrease in the attendance of spectators. We also find that NBA league games scheduled on the same day with MLB games negatively affected the attendance of spectators in between 34% (2008) and 81% (2013) of MLB games. While we do not find plausible explanations of why one of NHL or NBA games has more detrimental impact on the attendance of MLB games over years, we find that the substitution effects of both NHL and NBA games were prominent on specific years such as 2013, 2015 and 2018.

**Substitution Effects with Generic Factors:**

In this subsection, we estimate substitution effects that can be attributed to generic factors such as fan’s loyalty to MLB teams or fan’s preference of NHL or NBA games to MLB games. To this end, we create and present in Figure 3 a chart with two axes, the primary vertical axis (left side in blue color) to represent the number of MLB games conflicted NHL or NBA league games and the secondary vertical axis (right side in red color) to represent the proportion of conflicted MLB games that suffered from negative substitution effect.

We first note that the number of MLB games conflicted NHL or NBA league games varies significantly across MLB teams. For example, there are several MLB teams (e.g., BAL, CIN, KCR, SDP, and SFG) that have no games conflicted with NHL or NBA games mainly because cities hosting these MLB teams do not operate other professional leagues. While the majority of MLB teams (e.g., ARI, CLE, COL, and so on) had between 25 and 60 conflicted games with other leagues, few teams (e.g., BOS, CHC, CHW, and so on) had more than 60 conflicted games with other leagues.



**Figure 3:** Substitution Effect across Teams

In terms of the proportion of conflicted MLB games with negative substitution effect, most of teams that had more than 60 conflicted games experienced negative substitution effect (= decrease in the attendance) from 42% (BOS) to 78% (CHW) of conflicted MLB games. This great deviance of substitution effect indirectly insinuates that there are other MLB-team-specific factors to determine the magnitude of substitution effect. From limited data sets shown in Figure 3, among MLB teams with more than 60 conflicted games, MLB teams in the west division (e.g., LAA and LAD) suffered a relatively lighter substitution effect (50% to 51%), while several MLB teams in the central (e.g., CHC and CHW) and in the east division (e.g., NYM, NYY) experienced severe substitution effects (greater than 65%). However, it warrants follow-up studies to support the finding.

While substitution effects reported in Figure 3 present the aggregated substitution effects from both NHL and NBA league across MLB teams, substitution effects in Figure 4 show substitution effects for each league across MLB teams. Note that there are many MLB teams located in the cities that do not host NHL or NBA teams, which make them free from substitution effects by these alternative leagues. According to Figure 4, while several MLB teams (ATL, PIT, STL, and TBR) experienced substitution effect from only NHL league, several MLB teams (FLA, HOU, MIL, OAK, and SEA) experienced substitution effect from only NBA league.

Most MLB teams suffered substitution effects from bother NHL and NBA league with similar magnitude. However, interestingly, several MLB teams suffered a much severe substitution effect from NBA league than from NHL league: ARI (68% vs. 29%), COL (44% vs. 64%), and NYY (66% vs. 78%). In contrast, several other MLB teams suffered a much severe substitution effect from NHL league than from NBA league: BOS (49% vs. 37%), CHC (68% vs. 55%), MIA (80% vs. 46%), and MIN (88% vs. 58%).

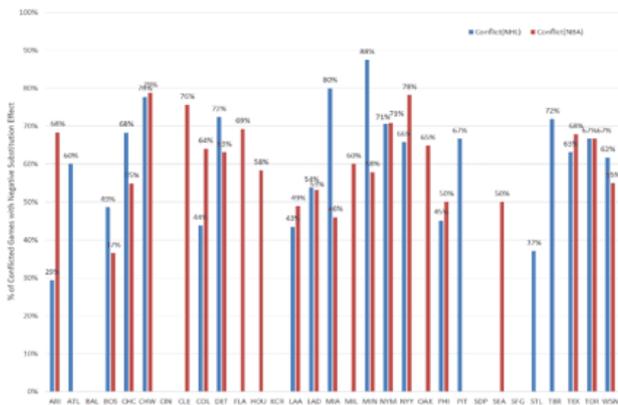


Figure 4: Substitution Effect of NHL and NBA Leagues across Teams

**Substitution Effects with Team Performance Factor:**

In this subsection, we try to estimate the substitution effects with team performance aggregation factors such as team places in the division and losing or winning streak in the past games. To this end, current 30 MLB teams are divided up evenly between the American League and National League and each of the leagues is divided into three divisions called the East, the Central, and the West. Therefore, MLB teams' standing in each division is always between 1 (best team) and 5 (worst team). However, the Houston Astros in the National League Central Division were reassigned to the American League West Division in 2013. Therefore, 23 records contain the sixth-place teams between 2008 and 2012 although we will not investigate them carefully due to the lack of records and asymmetry compared with other ranks in our data sets.

We speculate that MLB teams with lower standings will suffer from more severe substitutions effects than teams with higher standings. Our speculation is based on the findings in several studies (Lemke et al., 2010; Buraimo & Simmons, 2008;

Pawlowski & Anders, 2015) that audiences are more likely to attend to sports events when the chance of their home team winning the game increases.

We present a chart in Figure 5 that shows substitution effects of NHL or NBA leagues depending on MLB teams' standings in the division. According to Figure 5, the proportion of MLB games that have conflicted schedules with NHL games and lower attendance than average attendance is steadily increasing (i.e., from 57% to 59%, 61%, 61%, and 62%) as the standing of MLB teams in the division is worse (i.e., from 1st rank to 5th rank). Similarly, the proportion of MLB games that have conflicted schedules with NBA games and lower attendance than average attendance also shows an increasing trend (i.e., from 58% to 57%, 62%, 61%, and 63%) as the standing of MLB teams in the division is worse (i.e., from 1st rank to 5th rank).

When we consider conflicted games of both NHL and NBA with MLB games, we observe that MLB teams in high standings suffered less from negative substitution effects than MLB teams in low standings in the division. While MLB teams ranked at 6th suffered most from conflicted NBA games (i.e., 69% of MLB games with conflicted schedules lead to less attendance than average), they suffered least from conflicted NBA games (i.e., only 40% of MLB games lead to lower attendance than average).

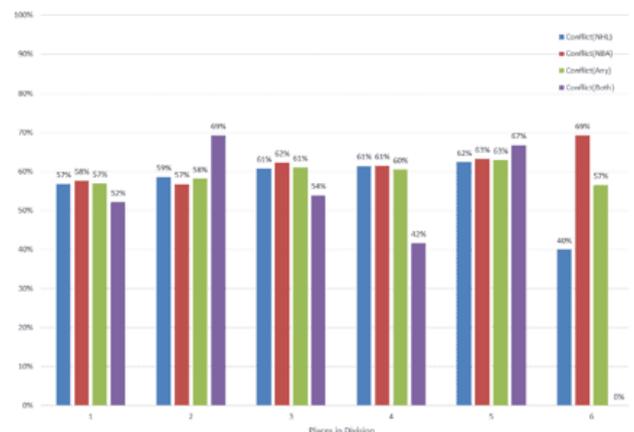
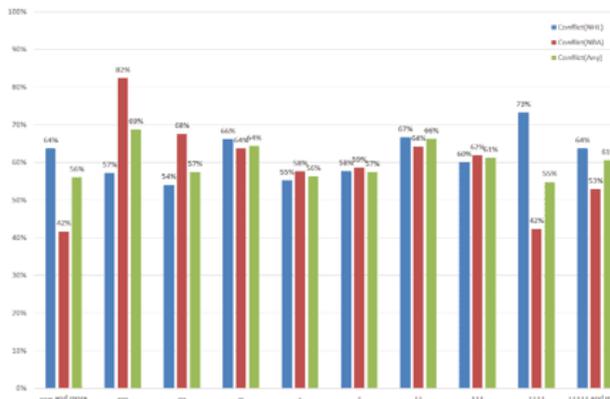


Figure 5: Substitution Effect with Team Places in Division

Another aggregation factor team performance we consider is the team's winning or losing streak. We anticipate that spectators are more likely to attend to sports events when the game between teams with rivalry relationships or similar ranks in the division is expected to be fun and exciting.

Based on this speculation, we posit that MLB teams in a long winning or losing streak suffer from more severely than other MLB teams. We summarize our findings in Figure 6.



**Figure 6:** Substitution Effect with Team's Winning or Losing Streak

In our data sets, several MLB teams experienced somewhere between the worst scenario of 9-games-losing streak (denoted as "-----") and the best scenario of 9-games-winning streak ("+++++++"). Therefore, we compute and present in Figure 6 the proportion of MLB games that have conflicted schedules with NHL games and lower attendance than average attendance for each value of team's winning or losing streak. Note that we combine the cases of winning or losing five or more games streak with conflicted schedules (25 records and 33 records out of 1,924 records, respectively) after considering the limited of records of such cases.

In Figure 6, we first note that in general, the proportion of MLB games with lower attendance than average attendance is relatively low when teams are in a five or more games of losing streak, which contradicts to our speculation that spectators will not be interested in watching games when their home teams lost several games in a row in the past, which should lead to the higher proportion of MLB games with lower attendance. However, it sharply increases as teams are in a four-game losing streak as we expected, reflecting that spectators show their deepest disappointments in their teams and avoid to attend MLB games in stadium. Since then, more teams suffer gradually less decrease in attendance as they perform better, reaching at the lowest at the status of one game loss.

We observe the symmetric trends as teams start to enjoy a case of winning streaks.

As they are in more games of winning streaks, the proportion of MLB games with lower attendance starts to increase, peaks at a four-game losing streak, and decreases at a five or more games of winning streak.

## 6. Conclusion

In this study, we estimate the direction and magnitude of substitution effects due to NHL or NBA games with schedule conflicts on the attendance demand of MLB games. In particular, we aggregate such substitution effects over three different factors such as temporal factors, team qualitative factor, and team performance factor.

Overall, we observe that substitution effects NHL and NBA games more significantly negatively impact the attendance on MLB games during the weekdays (in particular, Thursday and Wednesday) than during the weekend. We also observe that MLB teams in high standings suffer less from negative substitution effects than MLB teams in low standings in the division. In particular, when MLB teams are in a losing streak, spectators show their deepest disappointments in their teams and avoid to attend MLB games in stadium.

While our analysis in this paper provides general insights on the direction and magnitude of substitution effects in MLB franchises, its empirical implication on each MLB team is somewhat limited. Therefore, in future work, we intend to focus our analysis on the subjectively chosen MLB team so that the stakeholders of the MLB develop tailored marketing and promotion strategies to minimize the negative impact of competing sports on the attendance demand on MLB games.

## References

[1] Borland, J. & MacDonald, R. (2003). Demand for sport. *Oxford Review of Economic Policy*, 19(4), 478–502.

[2] Breiman, L. (1996). Bagging predictors. *Machine Learning*, 24, 123-140.

[3] Breiman, L.(2001).Random forests. *Machine Learning*, 45(1), 5-32.

[4] Buraimo, B. & Simmons, R. (2008). Do sports fans really value uncertainty of outcome? Evidence from the English Premier League. *International Journal of Sport Finance*, 3(3), 146–155.

- [5] Buraimo, B., Forrest, D., & Simmons, R. (2009). Insights for clubs from modelling match attendance in football. *Journal of Operational Research Society*, 60(2), 147–155.
- [6] Coates, D. & Humphreys, B.R. (2012). Game attendance and outcome uncertainty in the National Hockey League. *Journal of Sports Economics*, 13(4), 364–377.
- [7] Cox, A. (2018). Spectator demand, uncertainty of results, and public interest: Evidence from the English Premier League. *Journal of Sports Economics*, 19(1), 3–30.
- [8] Czarnitzki, D. & Stadtmann, G. (2002). Uncertainty of outcome versus reputation: Empirical evidence for the first German football division. *Empirical Economics*, 27(1), 101–112.
- [9] Dobson S. & Goddard, J. (2011). *The economics of football*. Cambridge, UK: Cambridge University Press.
- [10] Dubin, J.A. (2001). The demand for NFL football. In: *Empirical Studies in Applied Economics*. Springer, Boston, MA, pp. 31–49.
- [11] Forbes & Statista. (March 28, 2024). Major League Baseball total league revenue from 2001 to 2023. In: *Statista*. Retrieved January 13, 2025, from <https://www.statista.com/statistics/193466/total-league-revenue-of-the-mlb-since-2005/>
- [12] Forrest, D. & Simmons, R. (2002). Outcome uncertainty and attendance demand in sport: The case of English soccer. *Journal of the Royal Statistical Society: Series D*, 51(2), 229–241.
- [13] Forrest, D. & Simmons, R. (2006). New issues in attendance demand: The case of the English Football League. *Journal of Sports Economics*, 7(3), 247–266.
- [14] Freund, Y. & Schapire, R. E. (1995). A decision-theoretic generalization of on-line learning and an application to boosting. *Lecture Notes in Computer Science, Berlin, Heidelberg: Springer Berlin Heidelberg*, pp. 23–37.
- [15] Garcia, J. & Rodriguez, P. (2002). The determinants of football match attendance revisited: Empirical evidence from the Spanish Football League. *Journal of Sports Economics*, 3(1), 18–38.
- [16] Hart, R., Hutton, J., & Sharot, T. (1975). A statistical analysis of association football attendances. *Journal of the Royal Statistical Society: Series C*, 24(1), 17–27.
- [17] King, B. E., Rice, J. L., & Vaughan, J. (2018). Using machine learning to predict National Hockey League average home game attendance. *Journal of Prediction Markets*, 12(2), 85–98.
- [18] Knowles, G., Sherony, K., & Hauptert, M. (1992). The demand for Major League Baseball: A test of the uncertainty of outcome hypothesis. *The American Economist*, 36(2), 72–80.
- [19] 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.
- [20] Madalozzo, R. & Villar, R. (2009). Brazilian football: What brings fans to the game? *Journal of Sports Economics*, 10(6), 639–650.
- [21] Maszczyk, A., Zajac, A., & Rygula, I. (2011). A neural network model approach to athlete selection. *Sports Engineering*, 13(2), 83–93.
- [22] McCullagh, J. (2010). Data mining in sport: A neural network approach. *International Journal of Sports Science and Engineering*, 4(3), 131–138.
- [23] Martins, M.A. & Cro, S. (2018). The demand for football in Portugal: New insights on outcome uncertainty. *Journal of Sports Economics*, 19(4), 473–497.
- [24] Neale, W.C. (1964). The peculiar economics of professional sports: A contribution to the theory of the firm in sporting competition and in market competition. *The Quarterly Journal of Economics*, 78(1), 1–14.
- [25] Pang, Y. & Wang, F. (2024). Forecasting stadium attendance using machine learning models: A case of the National Football League. *Studia Sportiva*, 18(2), Publisher: Masaryk University Press.
- [26] Park, J., Cho, J., Gang, A.C., Lee, H.-W., & Pedersen, P.M. (2024). Machine learning prediction of factors affecting Major League Baseball (MLB) game attendance: Algorithm comparisons and macroeconomic factor of unemployment. *International Journal of Sports Marketing and Sponsorship*, 25(2), 382–395.

[27] Pawlowski, T. & Anders, C. (2012). Stadium attendance in German professional football—the (Un) Importance of uncertainty of outcome reconsidered. *Applied Economics Letters*, 19(16), 1553–1556.

[28] Reilly, B. (2015). The demand for league of Ireland football. *Economic and Social Review*, 46(4), 485–509.

[29] Rottenberg, S. (1956). The baseball players' labor market. *Journal of Political Economy*, 64(3), 242–258.

[30] Sahin, M. & Erol, R. (2017). A comparative study of neural networks and ANFIS for forecasting attendance rate of soccer games. *Mathematical and Computational Applications*, 22(4), 43–54

[31] Sahin, M. & Ucar, M. (2022). Prediction of sports attendance: A comparative analysis. Proceedings of the Institution of Mechanical Engineers, Part P: *Journal of Sports Engineering and Technology*, 236(2), 06-123.

[32] Serrano, R., Garcia-Bernal, J., Fernandez-Olmos, M., & Espitia-Escuer, M.A. (2015). Expected quality in european football attendance: Market value and uncertainty reconsidered. *Applied Economics Letters*, 22(13), 1051–1054.

[33] Strnad, D., Nerat, A., & Kohek, S. (2017). Neural network models for group behavior prediction: A case of soccer match attendance. *Neural Computing and Applications*, 28(2), 287–300.

[34] Varian, H. (2014). *Intermediate microeconomics*. (9<sup>th</sup> ed.). New York: W.W. Norton.

[35] Villa, G., Molina, I., & Fried, R. (2011). Modeling attendance at Spanish professional football league. *Journal of Applied Statistics*, 38(6), 1189–1206.

[36] Wallrafen, T., Nalbantis, G., & Pawlowski, T. (2022). Competition and fan substitution between professional sports leagues. *Review of Industrial Organization*, 61, 21–43.

Disclaimer / Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of Journals and/or the editor(s). Journals and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.