THE PENNSYLVANIA STATE UNIVERSITY
SCHREYER HONORS COLLEGE

DEPARTMENT OF ECONOMICS

AMERICA'S PAST TIME: AN EXPLORATION OF THE DECLINING ATTENDANCE AND VIEWERSHIP OF MAJOR LEAGUE BASEBALL

NIHAAR NARAYAN
SPRING 2020

A thesis
submitted in partial fulfillment
of the requirements
for baccalaureate degrees
in Economics & Computer Science
with honors in Economics

Reviewed and approved* by the following:

Peter Newberry
Assistant Professor of Economics
Thesis Supervisor

James R. Tybout
Professor of Economics
Honors Adviser

* Electronic approvals are on file.

ABSTRACT

This paper explores the trend of declining attendance and viewership of Major League Baseball, as well as the concern held by league officials and the media that the sport is losing the interest of the American audience. It investigates the effect that different in-game factors, which have been mentioned as possible contributors to the game's decline, have had on the league's attendance and viewership, roughly over the past decade. Specifically, this investigation uses multiple panel data regression analysis techniques, including random effects estimators and fixed effects estimators to glean the impact of increasing game time, slowing pace of play, and increasing Three True Outcome rate on attendance and viewership. Ultimately, this paper concludes that game time and pace of play are insignificant to attendance and viewership, while Three True Outcome rate negatively affects attendance. Within Three True Outcome rate, it is found that strikeout rate specifically negatively affects both attendance and TV ratings.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

I would like to thank my thesis supervisor Dr. Peter Newberry and my Economics 489 professor Dr. James Tybout for their support and guidance throughout my research.

Chapter 1

Introduction

Much has been made about the decline of the MLB and the sport of baseball over the last decade or so. The MLB has vocalized concerns about declining television viewership and attendance, and both for good reason. After experiencing decades of consistent growth in total attendance leading up to 2007, the league's attendance has steadily declined since then. The league's attendance between 1990 and 2007 increased relatively steadily from nearly 55 million to over 79 million (only interrupted by a steep drop and subsequent rebound in attendance due to the players' strike of 1994). Since then, attendance has dropped over 10 million, with the 2019 season only mustering up a total attendance of roughly 68.5 million (Baseball Reference). Additionally, the television ratings of the World Series have sharply decreased over the past few years, with individual World Series games in 2018 only generating viewership in similar quantities to that of regular season NFL games played around the same time (Bauder, 2018).

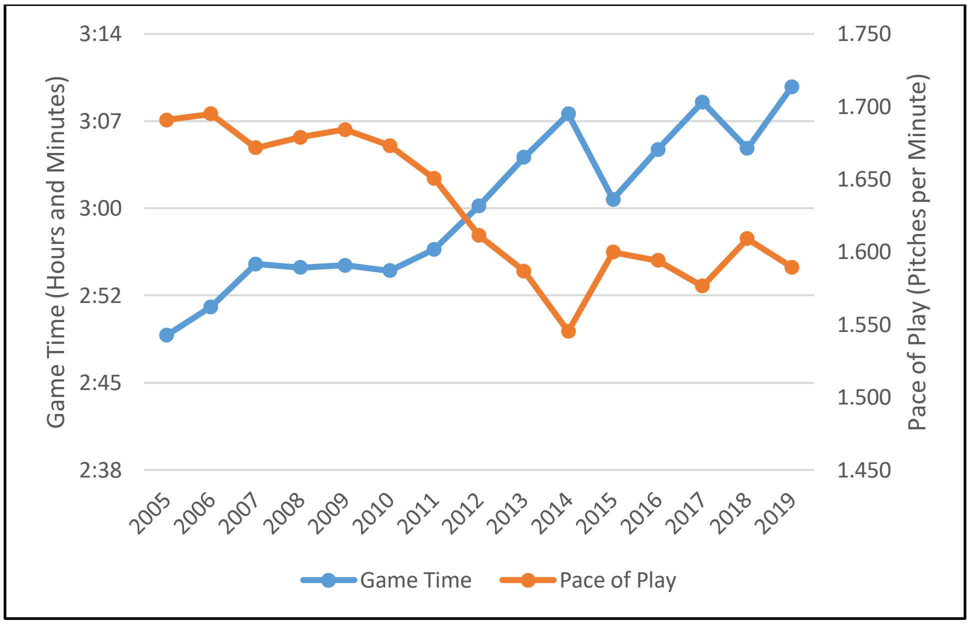
Despite both of these trends apparently signaling a growing lack of interest in the MLB, the league's revenues are higher than ever. League revenue came close to doubling over the span of 2007 to 2017, and while it seems that the MLB has generated much less interest among fans during recent years than the NFL or NBA, the rate of growth of the average MLB team is comparable to that of teams in the other two leagues. Over the last 22 years, as tracked by Forbes, the average MLB team valuation has increased by 11% every year, while this figure is 12% for the NFL and 13% for the NBA (Sykes 2019). While it is possible that concerns about declining interest might be slightly overblown, since the league seems to be as lucrative as ever, it could also be the case the league is able to maintain this growth because of its current broadcasting contracts. Although the league is locked into lucrative broadcasting contracts now, If the current trend of viewership continues, broadcasters may not be willing to pay the league as much for broadcasting rights moving forward, and major league teams may begin to see dips in their valuation.

Much of the dialogue surrounding the decline of the MLB's attendance and viewership has centered around the marketability of the game of baseball as it is currently played at the Major League level. This is where the game related factors that have been previously mentioned will make their way into this work. In particular, this exploration will focus on the length of baseball games, the pace of play, and style with which the game is played at the Major League level today.

1.1: Game Length and Pace of Play

To elaborate on these factors: MLB game length is relatively self-explanatory. Over the last 30 years, the average length of an MLB game has increased by 20 minutes but has increased by 10 of those minutes within the last decade. The average game length sits at 3 hours and 10 minutes in 2019, as opposed to hovering around 2 hours and 30 minutes in the 1970's (Baseball Reference). Pace of play is a similar concern for the league. While some of the increase in game time could be attributable to longer breaks between innings for advertising, much of it is attributable to a slowing down of the pace of play. This includes increased time between pitches, more frequent mound visits, more pitching changes, and more pitches per plate appearance. In 2019, both the number of pitches per plate appearance and the number of pitchers used by teams per game reached record highs. Particularly remarkable is the recent uptick in pitchers used per game. Over the last decade alone, this figure has increased from 3.87 to 4.41, an average increase of over .5 pitchers used per game (Baseball Reference). Both the lengthening of games and the slowing of pace can be seen in Figure 1 below.

Figure 1. MLB League Average Game Time and Pace of Play 2005-2019



Data from Baseball Reference

Pace of play has been such a concern for the league, that league's commissioner, Rob Manfred, has taken multiple steps to combat the slowing down of the game. Since he rose to the role of commissioner in 2015, Manfred has implemented a mound visit restriction, limiting teams to six mound visits per 9 innings of regulation (Kepner, 2018). The upcoming 2020 season will have a new rule requiring pitchers to face a minimum of three batters before they can be substituted out, which should effectively reduce the frequency of pitching changes, especially when it is not at all uncommon for relief pitchers nowadays to come out to face one or two batters (Bogage, 2019). Additionally, the league has kicked around the idea of a 20 second pitch clock, having implemented the rule at the minor league level as early as 2015 (Glaser, 2019).

Due to push back from the players' union, however, the rule's rollout at the major league level has been delayed to at least 2022 (Bogage, 2019).

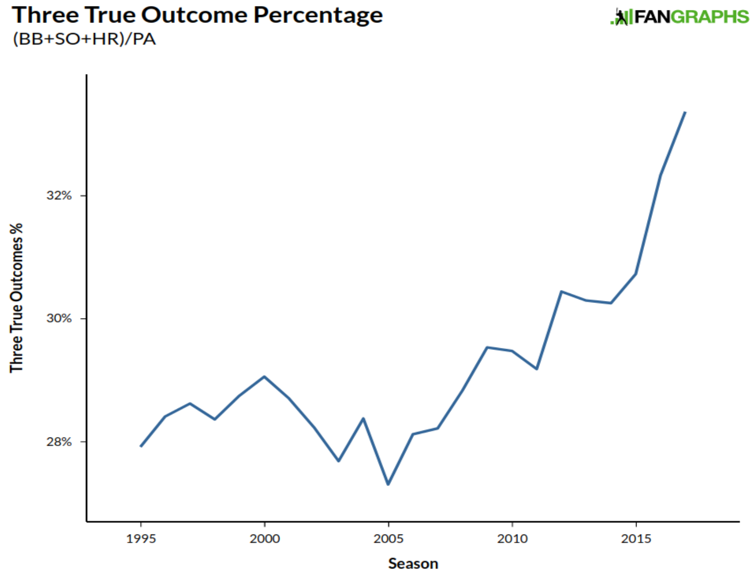
1.2: The Ever-Evolving Game and the Three True Outcomes

The last factor, the modern style of play, is a bit more nebulous, but itself likely plays a role in the increased game times and lower pace of play the MLB is experiencing. With data analytics and playing an increasingly large role in the decision-making of franchises, even at the managerial level, conventional "baseball knowledge" is being challenged, as is evident with the current trends of the game. Teams are employing defensive shifts more frequently, adapting to the individual tendencies of hitters. Starting pitchers are throwing far fewer innings, more relief pitchers are being used, and the traditional through line of starter, set-up man, closer is shifting, as traditional closers are pitching in high-leverage situations before the 9th inning. This breaking down of traditional roles for pitchers has led to more pitching changes, further contributing to inflating the length of Major League games. Additionally, hitters seemed to have changed their approach, as there has been a recent explosion in the frequency of what are called the "Three True Outcomes".

Comprised of the home run, the walk, and the strikeout, the Three True Outcomes are a concept that initially emerged from the online sabermetrics community as a bit of an inside joke, but quickly became a larger part of an analytics boom, revolutionizing the way that baseball is played. The Three True Outcomes are dubbed as such because they are, in theory, the only three possible outcomes of an at bat that are defense independent (Firstman, 2018). Whether a ball put in play results in the batter getting on base is up to the defense. As long as defenders are in fair territory, they can line up basically anywhere on the field (and they increasingly are, with the prominence of defensive shifts nowadays), so the success of any ball put in play depends on where the defenders are located, and whether they can make a play or not. From the pitcher's perspective, the strikeout is the only way to get an out without relying on defenders to get the out for you. From the hitter's perspective, drawing a walk in an at bat results in baserunner, independent of the defense. The home run, the last of the three outcomes and perhaps the ultimate outcome, is the only ball the batter can hit that is undefendable (except for inside-the-park home runs). While technically, a ball hit just barely over the fence is susceptible to being caught by an outfielder at the wall, the home run usually cannot be defended, regardless of the initial positioning of the outfielders. The "trueness" of these outcomes is well explained by Baseball Prospectus writer and co-founder Rany Jazayerli, who initially coined the term. "Together, the Three True Outcomes distill the game to its essence, the battle of pitcher against hitter, free from the distractions of the defense, the distortion of foot speed or the corruption of managerial tactics like the bunt and his wicked brother, the hit-and-run" (Baumann, 2017).

Whether entirely true or not, there is a prevalent narrative that the idea of the Three True Outcomes made its way into the minds of staff and players throughout the league, and that this development will drive fans away from the sport. In an article for The Ringer, Michael Baumann (2017) consulted major leaguers for their takes on this recent phenomenon. Former outfielder Carlos Beltrán not only noticed "more homers and more strikeouts," but a different attitude towards the strikeout coming from young hitters, remarking, "They feel like if they strike out, it's not a big deal." The 2017 leader in three true outcome rate (TTO%), Joey Gallo, described his approach to the plate, noting, "It seems like younger players, myself included, like going up to the plate with the intent to do damage. There's going to be some repercussions for that: You're probably going to strike out a little more, too" (Baumann, 2017).

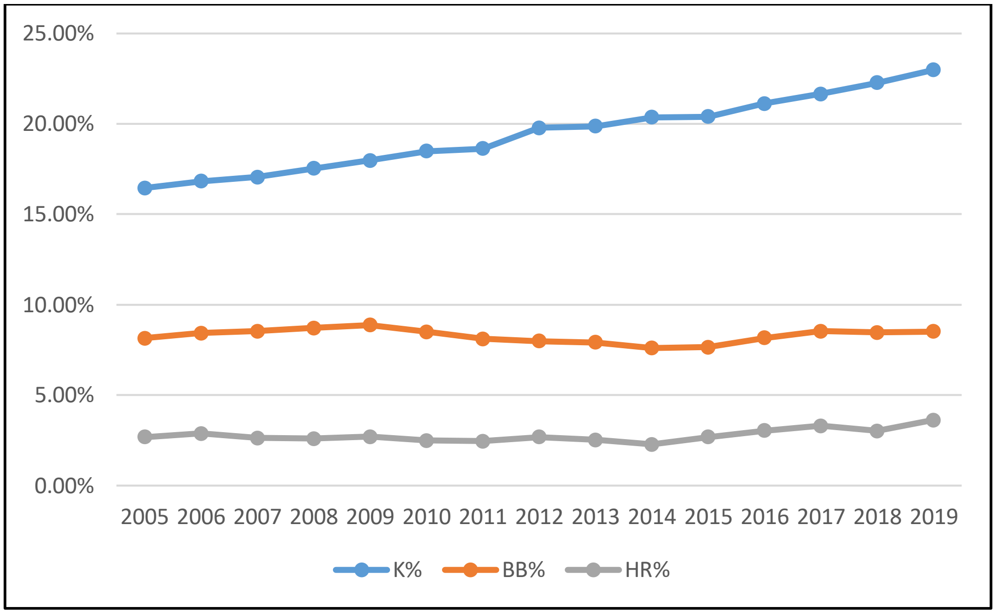
Figure 2. MLB League Average TTO% 1995-2017



Source: Edwards (2017)

The numbers certainly back up these observations. While the walk rate (BB%) has not changed much over the years, strikeout and home run rates (K% and HR%) have exploded, causing the league average TTO rate to shoot up over the last 15 or so years, as can be seen in Figure 1. The 2019 season saw more home runs and strikeouts than any year in the game's history (Baseball Reference).

Figure 3. MLB League Average K%, BB% and HR% 2005-2019



Data from Baseball Reference

For many, this is a worrisome trend, pushing the game of baseball towards an efficient extreme at the cost of making the game less exciting to watch. While home runs are plenty exciting, it is hard to say the same for walks and strikeout. With strikeouts climbing, less balls are being put in play than ever, which means less involvement from the defense, which means a lot more position players standing around on the field. As Baumann puts it, “There’s a point at which baseball might become so TTO-heavy, … that the sheer level of standing around will make the game boring” (Baumann, 2017).

This paper aims to use two separate panel regressions, having the dependent variables be attendance and television viewership. These are two metrics that are vital for determining the success, both monetarily and in popularity, of the league. Attendance data and viewership data will be collected for each team, over the course of several seasons, such that the unit of observation is one team-season. For each of these regressions, the explanatory variables of primary interest will be game time, pace of play, and three-true-outcome rate.

Chapter 2

Literature Review

Fortunately, there is a solid base of empirical work that covers determinants of attendance and viewership in Major League Baseball, as well as other professional sports organizations. This investigation is unique in the sense that the current body of literature largely focuses on finding determinants of attendance and viewership in the general sense, not so much in finding the determinants of a trend in attendance and viewership, just as this work will focus on the recent concern of declining attendance and viewership in Major League Baseball. Regardless, the methodologies used in the current body of literature act as great points of comparison to solidify the empirical approach this investigation will take.

2.1: Attendance Determinants in the MLB

In their study, *Major League Baseball Attendance: Long-Term Analysis Using Factor Models,* Ahn and Lee (2014) explored the determinants of attendance between Major League Baseball teams over the time period of 1904-2012. The findings of this study are particularly interesting. It finds that during the first half of this period, from 1904-1957, the only statistically significant determinant of a game's attendance is the winning percentage of the home team. On the other hand, the second half of this period, from 1958-2012, saw the baseball going audience become more nuanced in their discretion for attending games. Ahn and Lee find that, in addition to the winning percentage of the home team, the attendance of the game was also significantly impacted by outcome uncertainty, stadium size and quality, and play style.

Considering that the decline in viewership and attendance has been a recent phenomenon the league has been experiencing, the findings of the second period of Ahn and Lee’s study are pertinent. During this second period, it was mentioned that fans’ attendance was partially determined by “playing style.” They go on to elaborate that fans value good offensive play, indicating that there is a statistically significant preference for attendance as a team's batting average increases. To create the model used to arrive at their findings, they use a panel regression, regressing different determinants across teams and over the years between 1904 and 2012. Specifically, they used a fixed effects regression, controlling for time-invariant factors, as well as common factors across teams, or team-invariant factors. Considering the similar nature and goal of this investigation, as well as the identical panel structure of the data, a similar fixed effects regression would be appropriate. A pooled OLS estimator assumes that there are no time-invariant attributes that are unique to each team, which is almost certainly not the case. In the case of MLB teams, where attendance and viewership are heavily affected by mostly time-invariant factors such as market size, franchise pedigree, and fan loyalty, this assumption made by pooled OLS estimators means that it may not be a sufficient measure without the inclusion of a fixed effects estimator.

The limitation of Ahn and Lee’s study, largely as it pertains to being useful as a source for this investigation, is the lack of granularity when it comes to offensive and pitching metrics used to determine the effects of play style on fan attendance. Offensive performance was based entirely on batting average, while pitching performance was based only on strikeouts per game.

2.2: Television Viewership Determinants in the KBO

In another study, *Ex Ante and Ex Post Expectations of Outcome Uncertainty and Baseball Television Viewership*, Chang, Lee and Kang (2008) examined the ex ante and ex post determinants of television viewership of KBO League baseball games. The findings of this study state that, while ex ante expectations of outcome uncertainty have the most impact on viewership in the early innings of the game, ex post factors, such as in game outcome uncertainty and game quality became statistically significant towards the later innings. In addition to becoming more significant, their magnitudes of impact increased towards the end of the game, as well. Ex ante outcome uncertainty and ex ante perception of game quality still had large effects on viewership in the later innings, however.

To measure ex ante expectations of outcome uncertainty, this study found the difference in games back (lower difference is more uncertain) in the standings of both teams, while total games back between both teams (lower total is better) accounted for ex ante expectations of game quality. Ex post factors were measured by the flow of the game. Run differential measured outcome uncertainty as the game progressed, while total score and lead changes measured game quality. Their effect as determinants of viewership on some inning of a game between two teams was modeled both with a cross sectional data analysis, and a panel data analysis.

Chang, Lee and Kang's study is limited in relevance to this paper, simply because this paper will be focused very largely on ex ante expectations of game quality, and how overall trends of play have impacted these ex ante expectations and television viewership (as well as attendance). While difference in games back makes for a good ex ante indicator of a game's competitiveness, the worry around the league of declining ratings and attendance seems to have less to do with league parity or competitive balance, and more to do with the "watchability" of the game of baseball played at the Major League level (this may not be as much of a concern in the KBO). While the MLB is by no means an exemplary league in terms of parity, this is mostly be design. The league's lack of a hard salary cap has allowed bigger market teams, such as the New York Yankees, Boston Red Sox, and Los Angeles Dodgers, to vastly outspend their competition. This is not a recent development, however. Though it could absolutely be argued that improving league parity with a hard salary cap could make the league's competitive outcome more uncertain, and more likely to draw the interest of fans, the league has simply always been this way. If anything, the luxury tax (a "soft cap") implemented by the league in 1996 has been its most significant policy effort towards ensuring league parity. There may be some other factors related to how the game of baseball is played, and how this style of play has been trending recently, that have influenced ex ante expectations of game quality, and these factors are ultimately of interest to this paper.

2.3: Attendance Determinants in La Liga

In their study, *The Determinants of Football Match Attendance Revisited: Empirical Evidence From the Spanish Football League*, Garcia and Rodriguez (2002) explore the determinants of attendance in the Spanish Football League, or La Liga. The study ultimately found that factors representing ex ante quality of the game had significant positive determination on the attendance of a match. In addition, the study also found that the identity of the away team had a significant effect on attendance. If the away team was either FC Barcelona or Real Madrid, the two most popular teams in the league, attendance would increase. Even significant was when the away team was considered the "rival" of the home team. In fact, the study found that away team quality seems to have more of a significant effect on attendance than home team quality. This particular determinant, however, has little application to this exploration, as the effect of the away team is only a significant factor at the game level. Looking at a team's attendance at the season level, the large number of games played against many of the teams in the league should even out this effect across all teams in the MLB.

The limitation of this work is a microcosm of the limitations of the existing literature as it pertains to this paper. While there is a large body of existing literature that examines the determinants of attendance and viewership of professional sporting events, there is limited literature that focuses on the determinants of a trend in viewership or attendance. The recent decline in viewership and attendance in the MLB is a topic that has seen much speculation over the past decade, but little empirical study. While the goals of the body of literature may not match with the goals of this paper, the methodology used to model these determinants of attendance and viewership is surely something to learn from.

Chapter 3

Methodology

3.1: Variables and Data Collection

As was previously stated, the goal of this exploration is to investigate whether the MLB's recent decline in attendance and television viewership have been a product of some controversial game related changes the league gone through. These changes have been increased game length, slower pace of play, and higher TTO%, which will act as explanatory variables in the model. Game time is simply the average length of a game measured in minutes, per team, per year. Pace of play consists of the average pitch count divided by the average length of a game, per team, per year, resulting in a metric that measures a team's average pitches per minute. The three-true-outcome rate is a statistic consisting of the sum of walk rate, strikeout rate and home run rate, per team, per year. This is represented in a decimal, not a percent, so for a hypothetical interpretation where a TTO% of .280 increases by 1%, it increases to $.280 \times 1.01 = .2828$. The intuition is that game time and TTO% should have a negative impact on attendance and ratings, while pace of play should have a positive one.

Of course, these factors cannot be determining attendance and viewership in a vacuum, and it is unlikely they will even be the most significant cause of fluctuation in them. As was demonstrated in the research of Ahn and Lee (2014), teams will draw more attendance if they are winning. Simply put, the quality of a team is the prominent factor in determining how their attendance will fluctuate from year to year. To account for this, wins and division rank will be included as explanatory variables as well, both being good indicators of a team's performance in a given year. Additionally, fans likely carry over expectations from the success of the previous season when evaluating a team and considering whether to attend a game, especially when doing so early on in the current season. For this consideration, lag variables for a wins and division rank will be included, describing a team's wins and division placement from the season before. The intuition here is that wins and wins lag should have a positive effect on attendance and ratings, while division rank and rank lag should have a negative effect. Rank and rank lag should have a negative effect because an increase in the numerical value of a team's division rank from, say, 2 to 3, is actually the team dropping in the standings from second place to third place.

Fortunately, baseball statistics have been diligently compiled into organized databases by organizations such as Baseball Reference, so much of the relevant data for this exploration was available. All the data for the explanatory variables, including component data like total pitches thrown and faced per team, K%, BB% and HR%, was obtained from Baseball Reference. Attendance data was retrieved from ESPN's MLB Attendance Report and was collected from 2005-2019. There are two reasons for choosing this interval. The first is because it represents a period during all Major League franchises have remained in one city. Before 2005, the present-day Washington Nationals were the Montreal Expos, only moving to Washington for the 2005 season. The franchise's attendance was greatly impacted by this move, so this potential data skew was avoided by excluding attendance data from before the Nationals' move to Washington. The second reason for this interval is because it nearly coincides with the beginning of MLB's attendance decline. As was mentioned previously, the league's total attendance peaked in 2007, but has since fallen, so the interval of 2005-2019 captures this entire period, with two bonus years on the back end. Attendance data was collected in the form of average attendance per game, per team-season.

Regular season TV viewership ratings were retrieved from Forbes. Unfortunately, the regional ratings data available for each team was limited to the years 2014-2019. Data from 2014 was incomplete, as both the Los Angeles Dodgers and Houston Astros went through contract disputes with their regional broadcasters, so this year was excluded (Brown, 2014). That leaves the five-year window of 2015-2019 as the time frame for regressing regular season ratings. Moreover, ratings data was not available on the Toronto Blue Jays, presumably because they are in Canada. Because of this, the Blue Jays are excluded entirely from the ratings regression. The ratings data that was collected is local TV ratings for MLB teams' regional sports networks during primetime.

So that the data would be compatible with the necessary software, all the teams were given a team id, an integer representation between 1 and 30. All teams, with their corresponding team ids, are included in Appendix A.

3.2: The Model

Both the models for attendance and ratings are based on the two-way fixed effects model used in Ahn and Lee's (2014) study. Both the attendance and ratings models are described below.

$$\ln(y_{it}) = \beta_1 \ln(game\ time_{it}) + \beta_2 \ln(pace\ of\ play_{it}) + \beta_3 \ln(TTO\%_{it}) + \beta_4 \ln(wins_{it}) + \beta_5 \ln(division\ rank_{it}) + \beta_6 \ln(wins\ lag_{it}) + \beta_7 \ln(rank\ lag_{it}) + \alpha_{it} + u_{it}$$

- y_{it} represents either attendance or ratings (this model will be used for both regressions) for team i (= 1, … , 30) in year t (= 2005, … , 2019 for attendance; = 2015, … 2019 for ratings)

- α_{it} represents the unobserved time invariant heterogeneity, or "team effect"
- u_{it} represents random error

There are a few distinct features to this model, the first of which differentiates it from Ahn and Lee's model. Ahn and Lee used a two-way fixed effects model, accounting not only for the time invariant effect, but the team invariant effect, cause by factors common to all teams. The team invariant common factors are important in Ahn and Lee's model because of the 108-year window that the data occupies, from 1904-2012 (Ahn & Lee, 2014). Long term increases in attendance due to common factors across the league are inevitable in a period this large, so a team invariant fixed effect accounts for this. In this model, the time frame for attendance is only from 2005-2019, and for ratings its even smaller, so this is less of a concern.

A two-way fixed effects regression on the attendance data, using a least squares dummy variable (LSDV) approach with dummy variables for both teams and years, confirms this hypothesis.

Table 1. Dummy Variable P-Values for LSDV Two-Way Fixed Effects Regression

TEAM_ID	P > |t|
2	0.068
3	0.724
4	0.000
5	0.000
6	0.158
7	0.700
8	0.000
9	0.000
10	0.003
11	0.018
12	0.002
13	0.000
14	0.000
15	0.000
16	0.000
17	0.133
18	0.000
19	0.000
20	0.000
21	0.000
22	0.060
23	0.009
24	0.000
25	0.393
26	0.000
27	0.000
28	0.022
29	0.259
30	0.193

YEAR	P > |t|
2006	0.916
2007	0.235
2008	0.284
2009	0.321
2010	0.224
2011	0.360
2012	0.921
2013	0.477
2014	0.430
2015	0.695
2016	0.422
2017	0.356
2018	0.024
2019	0.010

In Table 1, above, cells highlighted in green demonstrate dummy variables that are significant at a 95% significance level (P < 0.05), while those cells in grey are not (P > 0.05). As was expected, there is significant heterogeneity among teams, with most team dummy variables being significant. Of the year dummy variables, however, only the most recent two were significant out of the 14. This indicates that the team-invariant effect in this dataset is quite limited, so including team-invariant common factors resulting in a two-way fixed effects model should be unnecessary.

The other feature of this model worth noting is that it is a log-log model (natural log used for both dependent and explanatory variables). One benefit of this regression technique is that it allows an otherwise linear model to fit non-linear trends. This is demonstrated with the algebraic manipulation below (model is abbreviated for demonstration).

$$\ln(y) = \beta_1 \ln(x_1) + \beta_2 \ln(x_2) + \beta_3 \ln(x_3)$$

$$e^{\ln(y)} = e^{\beta_1 \ln(x_1) + \beta_2 \ln(x_2) + \beta_3 \ln(x_3)}$$

$$e^{\ln(y)} = e^{\beta_1 \ln(x_1)} e^{\beta_2 \ln(x_2)} e^{\beta_3 \ln(x_3)}$$

$$y = x_1^{\beta_1} x_2^{\beta_2} x_3^{\beta_3}$$

Using a log-log approach also provides the benefits of meaningful interpretation of the model. In a log-log regression, the value of a coefficient β_i is the elasticity of the dependent variable with respect to the corresponding explanatory variable, x_i ($\beta_i = \frac{\%\Delta y}{\%\Delta x_i}$). In other terms, the interpretation of coefficient β_i can be stated as such: “If x_i increases by 1%, y increases by β_i%.” This interpretation is counterintuitive for division rank, and rank lag, as division rank can only take integer values between 1 and 5. A 1% change in rank is too small a change to reflect how division rank actually changes, so the coefficient for rank and rank lag will likely be very small. For the sake of consistency, however, these variables will still be logged. These variables, in addition to wins and wins lag, exist in this model because they likely have a significant impact on attendance and ratings, given their association with team quality. Their coefficients are not of particular interest compared to the coefficients of the game time, pace of play and TTO%.

3.3: Choosing a Regression Technique

While this model lends itself to a fixed effects estimator, there is another option that is suitable for this model, and that is the random effects estimator. Both the fixed effects estimator and the random effects estimator account for the unobserved time invariant heterogeneity, α_{it}. The distinction between the use cases for these two estimators is that the random effects estimators can only provide consistent estimates when this time invariant effect is not correlated with any of the explanatory variables. The requirement for a random effects estimator to be viable, for explanatory variable x_1 and unobserved heterogeneity α is

$$Cov(\alpha, x_i) = 0 \mid i \in [1, n],$$

where n is the number of explanatory variables. When this condition is met, random effects is actually a better estimator than fixed effects and will generally estimate with less standard error.

The Hausman test performs a hypothesis test on the null hypothesis that the above condition is true. This test calculates a test statistic (shown below), called the Hausman statistic (H) using a chi-square distribution with n degrees of freedom, where n is the number of explanatory variables. $\hat{\beta}_{FE}$ represents the fixed effects estimators estimate of coefficient β, while $\hat{\beta}_{RE}$ is same estimate made by the random effects estimator.

$$H = \frac{(\hat{\beta}_{FE} - \hat{\beta}_{RE})^2}{Var(\hat{\beta}_{FE}) - Var(\hat{\beta}_{RE})} \sim \chi^2(7)$$

If the p-value for this test statistic is < 0.05, we reject the null hypothesis at a 95% significance level, and fixed effects is the only consistent estimator. If the p-value for this test statistic is $>$ 0.05, then we fail to reject the null hypothesis at a 95% significance level, and random effects is preferred, as the more optimal estimator.

For the attendance data, the result of the Hausman test resulted in a test statistic $W = 15.83$, and a p-value $Prob > W = 0.0267$. This p-value is indeed less than 0.05, so we reject the null for the attendance data. For the ratings data, the results look surprisingly different. It resulted in a test statistic $W = 5.33$, and a p-value $Prob > W = 0.6197$. This p-value is far greater than 0.05, so we fail to reject the null hypothesis that $Cov(\alpha, x_i) = 0 \mid i \in [1, n]$. Based on these tests, a fixed effects estimator is more suitable for the attendance data, while a random effects estimator is more suitable for the ratings data.

Chapter 4

Results

4.1: Attendance Results

Table 2. Results of Fixed Effects Estimator for ln(Attendance)

$\boldsymbol{ln}(\boldsymbol{Attendance})$	Coefficient	Standard Error	P>|t|
$\ln(game\ time)$	0.4500616	0.3390826	0.185
$\ln(pace\ of\ play)$	0.3394409	0.3540303	0.338
$\ln(TTO\%)$	-0.3876348	0.0785120	0.000
$\ln(wins)$	0.3601964	0.0745852	0.000
$\ln(division\ rank)$	-0.0100427	0.0183208	0.584
$\ln(wins\ lag)$	0.3148349	0.0764828	0.000
$\ln(rank\ lag)$	-0.0376289	0.0184917	0.042

Of the seven explanatory variables in this regression, four of them are significant at a 95% significance level: TTO%, wins, wins lag and rank lag. Wins and wins lag being significant are no surprise. As indicated by Ahn and Lee (2014), a team that is winning is a team that can fill the seats with fans. Both the wins coefficient and the wins lag coefficient are similar, with the wins coefficient being slightly higher. The wins coefficient indicates, that if a team wins 1% more games, its average attendance per game will increase by 0.3601964%. The wins lag coefficient indicates, that if a team has won 1% more games the previous year, its average attendance per game would have increased by 0.3148349% this year.

Interestingly, these figures are quite a bit lower than those found in Ahn and Lee’s results. Based on their results, they found that if a team increases its win percentage by 1%, it will see its attendance rise by about 1.9%, during the period from 1958-2012 (Ahn & Lee, 2014).

Part of this is attributable to the fact that winning 1% more games is different than increasing win percentage by 1 percentage point. For example, an 81-win team with a 50% win percentage would end up with 81.81 wins if it were to win 1% more games, but with 82.62 wins if it were to increase its win percentage by one percentage point. Each percentage point in a team's winning percentage represents 1.62 wins in a complete 162 game season, and 1 percent increase to wins will always be a smaller increase than that. Even so, the average MLB team wins 81 games, and at that level, a 1 percentage point increase in win percentage is only twice as much as a 1 percent increase to wins. This model's coefficient for wins is roughly 5 times smaller than Ahn and Lee's 1.9, so there should be another reason behind this.

The intuitive interpretation of this is that fan demand for attendance has grown inelastic over the years, at least with respect to wins. The time frame for the later period in Ahn and Lee's model is 1958-2012, while this attendance model only spans from 2005-2019 (Ahn & Lee, 2014). Perhaps the improvement of ballpark facilities have made attending a baseball game more appealing to the casual fan, regardless of the quality of the team. On the other hand, perhaps teams are having a hard time attracting new fans, so even teams that are having good seasons aren't pulling in as many new fans as they used to. There are a number of different plausible stories that could explain why demand for attendance may have become more inelastic. This discrepancy may be due to error in the model, however, and it may be instead that the presence of some other explanatory variables is contributing to the weakening of the effect of wins.

In addition to wins and wins lag, rank lag is significant, indicating if the numerical value of a team's rank were to increase by 1%, the team's attendance would decrease by 0.0376289%. The negative coefficient for rank lag was expected, but it is interesting that rank lag is significant, while division rank is not. This could be because fans carry over a lot of expectation from a team's placement in the standings from the previous season, and it takes a longer time these expectations to adjust to the team's current placement in the standings.

The variable of interest here is TTO%, which was highly significant with a p-value of 0, and the expected negative coefficient. The regression predicts that a 1% increase in TTO% (recall; a 1% increase in the decimal value, not a percentage point increase) will decrease attendance by 0.3876348. The magnitude of this coefficient is quite substantial, considering a 1% increase in a typical value for TTO%, such as .300, would only result in an increase to .303. Ahn and Lee's findings return to mind here as well. Their findings stated that modern fans (1958-2012), had a significant preference for offense; specifically, higher batting average (Ahn & Lee, 2014). TTO% is comprised of HR% and BB% both of which are offensive components, and the former of which is one of the most exciting and iconic parts of baseball. Even though the larger narrative on the increase in TTO% has been negative, it follows that if fans prefer offense, that they would respond well to more home runs. While home runs have increased to record highs over the past few years, so have strikeouts. On top of that, home runs are simply infrequent compared to strikeouts, and even walks (Average K%, BB%, HR% from 2005-2019: .1943, .0828, .0278). The increase in TTO% over the last several years has been largely driven by an increase in strikeouts. Given the philosophies of younger hitters, it appears that the increase in home runs has come at the cost of an increase in strikeouts. Ultimately, it seems that the increase in TTO% has had a net negative impact on the "offensive quality" of the game, leading to it adversely affecting attendance.

Surprisingly, neither game time nor pace of play were significant in this regression. Game time's coefficient indicates that longer games should positively impact attendance, which is unexpected, but its lack statistical insignificance makes indicates this is a potentially dubious interpretation.

4.2: Ratings Results

Table 3. Results of Random Effects Estimator for ln(Ratings)

$\boldsymbol{ln(Ratings)}$	Coefficient	Standard Error	P>|z|
$\ln(game\ time)$	0.4305920	1.0148480	0.671
$\ln(pace\ of\ play)$	0.3891562	1.5069710	0.796
$\ln(TTO\%)$	-0.4684917	0.2972703	0.115
$\ln(wins)$	1.3674840	0.2060073	0.000
$\ln(division\ rank)$	-0.0150700	0.0570527	0.792
$\ln(wins\ lag)$	0.3784124	0.2237197	0.091
$\ln(rank\ lag)$	-0.0788680	0.0564927	0.163

Of the explanatory variables, only wins is a statistically significant determinant of TV ratings at a 95% significance level. The other three variables that were significant in the previous attendance regression have relatively low p-values but are not quite significant. This could be a product of the limited time window of this regression, as the data was limited to the years 2015-2019. Perhaps with more data, these factors would have a clearer effect on ratings.

Once again, the statistical significance and sign of wins is unsurprising, but the magnitude of the coefficient is interesting, as it is quite a bit higher than the coefficient of wins in the attendance regression. The ratings model predicts that a 1% increase in wins will result in a 1.3674840% increase in ratings, indicating that fans demand for televised baseball is more elastic than their demand to attend baseball games, with respect to wins. This makes some intuitive sense. Attending a baseball game may be more win inelastic because of the other benefits it provides as an attraction. Additionally, going to a baseball game takes some level of commitment, both of effort and of money, so there may be other more personal considerations fans consider when attending a baseball game. Watching a game on TV, on the other hand, requires little commitment to do, and is just as easy to stop doing. Because of this, fans may be quicker to stop watching games if their team is not winning.

While pace of play and game time appear to be statistically insignificant factors for both attendance and ratings, TTO% is promising. It is significant when regressed on attendance, and nearly significant when regressed on ratings. It is however, as previously discussed, a bit of a mixed bag statistically. K%, which makes up the majority of TTO%, should negatively affect attendance and ratings, while HR%, which makes up a small portion of TTO%, should positively affect them. Splitting up TTO% into its components should clarify which aspects of TTO% have sway over attendance and ratings. The new model is below.

$$\ln(y_{it}) = \beta_1 \ln(game\ time_{it}) + \beta_2 \ln(pace\ of\ play_{it}) + \beta_3 \ln(K\%_{it}) + \beta_4 \ln(HR\%_{it}) + \beta_5 \ln(BB\%_{it}) + \beta_6 \ln(wins_{it}) + \beta_7 \ln(division\ rank_{it}) + \beta_8 \ln(wins\ lag_{it}) + \beta_9 \ln(rank\ lag_{it}) + \alpha_{it} + u_{it}$$

4.3: Attendance Results w/ TTO% split

Table 4. Results of Fixed Effects Estimator for ln(Attendance) w/ TTO% split

$\boldsymbol{ln(Attendance)}$	Coefficient	Standard Error	P>|t|
$\ln(game\ time)$	-0.0574430	0.3684575	0.876
$\ln(pace\ of\ play)$	-0.5559762	0.4412973	0.208
$\ln(K\%)$	-0.3677794	0.0619625	0.000
$\ln(HR\%)$	0.0496125	0.0360490	0.169
$\ln(BB\%)$	0.0634605	0.0731348	0.386
$\ln(wins)$	0.2790204	0.0777363	0.000
$\ln(division\ rank)$	-0.0064706	0.0181263	0.721
$\ln(wins\ lag)$	0.2970357	0.0758743	0.000
$\ln(rank\ lag)$	-0.0356147	0.0183320	0.053

Outside of TTO%, not much has changed, but inside of TTO%, K% seems to have by far the most significant effect. It is the only statistically significant factor of the Three True Outcomes, and its coefficient is all but as large in magnitude as the coefficient of TTO% in the original attendance model. This model predicts that a 1% increase in the decimal value of K% will lead to attendance decreasing by 0.3677794%. This reinforces the idea that the negative impact of the trend of increasing TTO% is at the very least, largely attributable to increasing K%.

4.4: Ratings Results w/ TTO% split

Table 5. Results of Random Effects Estimator for ln(Ratings) w/ TTO% split

$\boldsymbol{ln(Ratings)}$	Coefficient	Standard Error	P>|z|
$\ln(game\ time)$	-0.1264323	1.1048840	0.909
$\ln(pace\ of\ play)$	0.4831634	1.5134220	0.750
$\ln(K\%)$	-0.5834825	0.2708558	0.031
$\ln(HR\%)$	0.2355827	0.1489255	0.114
$\ln(BB\%)$	-0.1737619	0.2311783	0.452
$\ln(wins)$	1.2960220	0.2146432	0.000
$\ln(division\ rank)$	0.0012757	0.0569410	0.982
$\ln(wins\ lag)$	0.2767584	0.2295139	0.228
$\ln(rank\ lag)$	-0.0977720	0.0568860	0.086

After splitting up the TTO% components for the ratings model, both wins and K% are statistically significant at a 95% significance level. The coefficient for wins is similar to before. The interpretation for K% indicates that a 1% increase in the decimal value of K% will lead to a 0.5834825% decrease in ratings. Much how in the original attendance and ratings models, TV ratings were more win elastic than attendance was, ratings are more K% elastic than attendance. To put it in another way, fans watching at home would be more likely to switch channels if the home team is swinging and missing than fans are to sell their tickets.

Chapter 5

Conclusion

The objective of this exploration was to investigate the validity of the different concerns regarding the MLB and its declining attendance and television viewership. These concerns were increasing game times, slower pace of play, and increasing TTO%. While there is certainly sound intuition behind the thought that a slower and longer game would be considered boring by many fans, causing bleachers and couches to go unfilled, these factors were not statistically significant in determining either attendance or ratings. Aside from wins, the other significant factor of note was TTO%, and specifically K%. An increasing K% was estimated to have an adverse effect on both attendance and ratings. These findings, as well as the findings that wins are a significant factor in estimating attendance and ratings, are consistent with the findings of Ahn & Lee (2014). They found that not only do the modern baseball fans expect their team to win, but they also have a significant preference for offensive play. It then stands to reason that fans would have an aversion to strikeouts.

While the league seems quite concerned with shortening games and speeding up pace, its quite possible that it should be worried about what the game is becoming. The league is heading into new territory with the march of the Three True Outcomes, but it may not be the direction that many fans want to see the sport go.

Appendix A

MLB Teams with Associated Team IDs

TEAM	TEAM_ID
Arizona Diamondbacks	1
Atlanta Braves	2
Baltimore Orioles	3
Boston Red Sox	4
Chicago Cubs	5
Chicago White Sox	6
Cincinnati Reds	7
Cleveland Indians	8
Colorado Rockies	9
Detroit Tigers	10
Houston Astros	11
Kansas City Royals	12
Los Angeles Angels	13
Los Angeles Dodgers	14
Miami Marlins	15
Milwaukee Brewers	16
Minnesota Twins	17
New York Mets	18
New York Yankees	19
Oakland Athletics	20
Philadelphia Phillies	21
Pittsburgh Pirates	22
San Diego Padres	23
San Francisco Giants	24
Seattle Mariners	25
St. Louis Cardinals	26
Tampa Bay Rays	27
Texas Rangers	28
Toronto Blue Jays	29
Washington Nationals	30

BIBLIOGRAPHY

Ahn, S. C., & Lee, Y. H. (2014). Major League Baseball Attendance: Long-Term Analysis Using Factor Models. Journal of Sports Economics, 15(5), 451–477. doi: 10.1177/1527002514535171

Bauder, D. (2018, October 31). World Series ratings sharply down from last year. Retrieved from https://apnews.com/754cd23aba144d998dd76b4e217d62ba

Baumann, M. (2017, August 7). The End of Baseball As We Know It. Retrieved from https://www.theringer.com/2017/8/7/16108098/the-end-of-baseball-as-we-know-it

Bogage, J. (2019, February 27). MLB reportedly offers to postpone pitch clock until 2022. Retrieved from https://www.washingtonpost.com/sports/2019/02/27/mlb-reportedly-offers-postpone-pitch-clock-until/?arc404=true

Brisbee, G. (2018, July 12). Why baseball games are so damned long. Retrieved from https://www.sbnation.com/a/mlb-2017-season-preview/game-length

Brown, M. (2014, October 3). MLB Completely Dominated Local Prime Time TV During The 2014 Season [UPDATED]. Retrieved from https://www.forbes.com/sites/maurybrown/2014/10/03/mlb-completely-dominated-local-prime-time-tv-during-the-2014-season/#6ccea03a45d4

Brown, M. (2018, October 3). Why MLB Attendance Dropped Below 70 Million For The First Time In 15 Years. Retrieved from https://www.forbes.com/sites/maurybrown/2018/10/03/how-mlb-attendance-dropped-below-70-million-for-first-time-in-15-years/#29c8de3161bf

Brown, M. (2019, October 15). 2019 MLB Regional TV Ratings In Prime Time Remain Solid. Retrieved November 1, 2019, from https://www.forbes.com/sites/maurybrown/2019/10/15/2019-mlb-regional-tv-ratings-in-prime-time-remain-solid/#85a80653f89e.

Chung, J., Lee, Y. H., & Kang, J.-H. (2016). Ex Ante and Ex Post Expectations of Outcome Uncertainty and Baseball Television Viewership. Journal of Sports Economics, 17(8), 790–812. https://doi.org/10.1177/1527002514551002

Edwards, C. (2017, August 11). Young Players Are Leading the Rise in Three True Outcomes. Retrieved from https://blogs.fangraphs.com/young-players-are-leading-the-rise-in-three-true-outcomes/#more-261405

Firstman, D. B. (2018). The Growth of 'Three True Outcomes': From Usenet Joke to Baseball Flashpoint. *Baseball Research Journal*, *47*(1).

Forbes Magazine. (n.d.). The Business Of Baseball. Retrieved November 1, 2019, from https://www.forbes.com/mlb-valuations/list/.

García, J., & Rodríguez, P. (2002). The Determinants of Football Match Attendance Revisited: Empirical Evidence From the Spanish Football League. Journal of Sports Economics, 3(1), 18–38. doi: 10.1177/152700250200300100З

Glaser, K. (2019, February 22). MLB Announces 20-Second Pitch Clock For 2019 Spring Training Games. Retrieved from https://www.baseballamerica.com/stories/mlb-announces-20-second-pitch-clock-for-spring-training-games/

Kepner, T. (2018, February 19). M.L.B. Puts a Limit on Mound Visits Per Game. Retrieved from https://www.nytimes.com/2018/02/19/sports/baseball/mlb-mound-visits.html

Major League Baseball Miscellaneous Year-by-Year Averages and Totals. (n.d.). Retrieved from https://www.baseball-reference.com/leagues/MLB/misc.shtml

MLB Attendance Report - 2019. (n.d.). Retrieved November 1, 2019, from http://proxy.espn.com/mlb/attendance?sort=homePct.

Sykes, M. (2019, April 12). The ever-increasing value of MLB teams. Retrieved from https://www.axios.com/major-league-baseball-team-value-a8454505-cc7d-4bb3-9406-d4f11cce7f80.html

ACADEMIC VITA

NIHAAR NARAYAN

EDUCATION/SKILLS

Penn State University Schreyer Honors College, University Park *Graduation: May 2020*

- B.S. in Computer Science, Economics
- **Relevant Coursework:** Operating Systems, Data Structures & Algorithms, Theory of Computation, Programming Language Concepts, Object Oriented with Web Applications, Discrete Math for Computer Science, Honors Calculus III
- **Programming Languages/Libraries:** Python, Java , JavaScript, C, C#, C++, CSS, Golang, Handlebars.js, Knockout.js
- **Other Technologies:** Kubernetes, Unity Game Engine, Autodesk MotionBuilder, Autodesk Maya

WORK EXPERIENCE

Software Development Intern at Appian Corporation ***June – August 2019***

- Software development intern on the Kubernetes Monitoring & Logging team at Appian headquarters. Managed and monitored Kubernetes deployments in the cloud using AWS, Helm charts, Prometheus and Grafana. Wrote script to automate release signing during my team's downstream deployment process. Created a generative testing tool for Appian's Expression Rule Designer during an intern hackathon. Regularly presented new features to tribe management at retrospective meetings.

Software Development Intern at Analytical Graphics Inc. (AGI) ***May - August 2018***

- Software development intern in the Exton Headquarters of AGI. Worked on the enterprise platform team, and was an active participant in their AGILE development process and release cycle. Implemented a custom geocoder front-end, which was able to take an English description of a place and focus on the corresponding location on a 3D globe. Also implemented new Firefox context menu workflow for STIG compliance.

Software Development Intern at Bentley Systems ***May - August 2017***

- Software development intern in the Exton Headquarters of Bentley Systems. Implemented a new file naming workflow for Bentley's AECOsim Building Designer to meet the Singapore Government's Building and Construction Authority Standards.

RESEARCH

Summer Research Intern: University of Pennsylvania ***June - August 2014, 2015***

- Animated and programmed character models to create a behavioral simulation of the Reading Terminal Market in the Unity game engine. Added more accurate motion detection to a VR baseball hitting simulator for the Xbox Kinect.

EXTRACURRICULAR/ACHIEVEMENTS

- **WKPS Radio Co-host:** Co-host of the Suburban Sound on WKPS, The LION 90.7, Fridays 7 - 9 PM
- **President's Freshman Award, SHC Academic Excellence Scholarship, National Merit Finalist**

LINKS

- **LinkedIn:** https://www.linkedin.com/in/nihaar-n-945235112/