THE DETERMINANTS OF MAJOR LEAGUE BASEBALL ATTENDANCE

A THESIS

Presented to

The Faculty of the Department of Economics and Business

The Colorado College

In Partial Fulfillment of the Requirements for the Degree

Bachelor of Arts

By

Scott Thauwald
May / 2008
THE DETERMINANTS OF MAJOR LEAGUE BASEBALL ATTENDANCE

Scott Thauwald

May, 2008

Economics

Abstract

The purpose of this study is to examine different factors that affect Major League Baseball (MLB) attendance. In doing so, this study uses one dependent variable and twelve independent variables. Data for the study was gathered for five years, starting in 2003 and ending in 2007. A regression analysis was used to analyze the data and statistics. The results found seven variables to be significant. The significant variables include payroll, regular season wins, average ticket price, facility age, substitute teams, income, and arena capacity. Of the seven variables, payroll and regular season wins were found to be the most influential.

KEYWORDS: (Attendance, Baseball, Major League Baseball)
ON MY HONOR, I HAVE NEITHER GIVEN NOR RECEIVED UNAUTHORIZED AID ON THIS THESIS

Signature

Scott Thurull
# TABLE OF CONTENTS

ABSTRACT .................................................................................................................. iii  
I INTRODUCTION ......................................................................................................... 1  
II LITERATURE REVIEW AND MODEL DEVELOPMENT ........................................... 6  
   Sports Attendance................................................................................................. 7  
      Winning Record.................................................................................................. 7  
      Ticket Price...................................................................................................... 9  
      Payroll............................................................................................................... 10  
      Facility Age..................................................................................................... 11  
      Geographical Location..................................................................................... 13  
      Substitute Teams.............................................................................................. 14  
      Star Players..................................................................................................... 15  
      Homeruns....................................................................................................... 16  
      Strikes............................................................................................................ 17  
      Promotions..................................................................................................... 18  
      Playoff.......................................................................................................... 19  
      City Population............................................................................................... 20  
      Competitive Balance....................................................................................... 21  
      Income........................................................................................................... 22  
      Interleague Play.............................................................................................. 23  
      Summary......................................................................................................... 25  
III METHODOLOGY .................................................................................................... 27  
   Dependent Variable............................................................................................ 27  
   Control Variables................................................................................................. 28  
   Independent Variables......................................................................................... 28  
   Summary............................................................................................................ 31  
IV RESULTS ................................................................................................................ 32  
   Model Equations.................................................................................................. 35  
   Regression Analysis Results............................................................................... 37  
V CONCLUSION .......................................................................................................... 46  
   Possible Limitations and Further Research....................................................... 49  
SOURCES CONSULTED ............................................................................................. 52
LIST OF TABLES

1.1 2007 MLB Teams................................................................. 4

2.1 Past Research on Factors Affecting Attendance in Sport Leagues .......... 24

3.1 Variables Summary............................................................. 31

4.1 Summary Statistics............................................................. 32

4.2 Correlation Matrix.............................................................. 33

4.3 Regression Analysis Results.................................................. 37

4.4 Regression Analysis Results.................................................. 43
LIST OF FIGURES

1.1 MLB Attendance............................................................................................. 2

4.1 Model Equation: Control Variables.................................................................... 35

4.2 Model Equation: Base Model............................................................................ 35

4.3 Model Equation: Significant Variables............................................................... 35

4.4 Model Equation: Base Minus Payroll................................................................. 36

4.5 Model Equation: Significant Base Minus Payroll............................................. 36
CHAPTER I
INTRODUCTION

When the 2007 Major League Baseball (MLB) season came to an end, teams across the country were trying everything they could do to fill as many seats as possible. Whether it’s the Rockies, who clinched a playoff spot, or the Twins teams want to end a season by generating money through ticket sales. But determining what brings fans to games is a problem that a lot of teams face. Is it the fact that a team is good, the chance to see a star player, or is it because of a stadium?

This thesis examines the different factors that affect attendance at a Major League Baseball game. In doing so, this thesis will gather data from the 2003 season all the way to the 2007 season and will then come to a conclusion on which factors affect MLB attendance.

The game of baseball has been known as America’s pastime since the late 19th century. Many kids throughout America spend their nights and weekends throwing the ball around thinking they are going to be the next Alex Rodriguez. For many baseball players it is their dream to play Major League Baseball, and that’s one of the reasons why the MLB has such a good following.
The great game of baseball today is more popular than ever with millions of diehard fans watching and supporting their favorite teams. Figure 1.1 shows the total attendance number from 1997 to 2006. MLB attendance has continued to increase over the years, even with a slight dip in 2002 and 2003. In 2006, there were a total of 77,532,108 fans who attended a MLB game, many being families with kids. Baseball will always be a great family affair but over the years Major League Baseball has become more of a business than a game. With payrolls increasing and winning more important than ever, teams need to put people in the stands to generate money. Teams know the more fans, the more money they will make, therefore more money to spend. If teams have the luxury to spend massive amounts of money on players, they will likely be a
more competitive team. Because attendance is so important teams are trying to do everything they can to get fans in the seats. But what is it exactly that brings fans to games? In the 2006 season, the New York Yankees averaged 51,000 fans a game whereas the Florida Marlins averaged 13,500. Why did the Yankees average so many more fans than the Marlins when both teams had a shot at making the playoffs? Was the weather a factor considering Florida becomes very hot during the baseball season or was it something else? The purpose of this thesis is to answer many of these questions about MLB attendance and help understand what variables affect getting people to a MLB game.

What is the MLB?

The MLB stands for Major League Baseball and is one of the four major sports in America. The MLB is the highest level of baseball play in North America and the league consists of thirty teams spread across the country. The MLB is divided into two leagues, the American League and the National League. In each league there are three divisions, East, Central, and West. There are anywhere from four to six teams in each division. Teams for the 2007 season look as follows:
TABLE 1.1

2007 MLB Teams

<table>
<thead>
<tr>
<th>Division</th>
<th>National League (NL)</th>
<th>American League (AL)</th>
</tr>
</thead>
<tbody>
<tr>
<td>*Central</td>
<td>Chicago Cubs, Cincinnati Reds, Houston Astros, Milwaukee Brewers, Pittsburgh Pirates, St. Louis Cardinals</td>
<td>Chicago White Sox, Cleveland Indians, Detroit Tigers, Kansas City Royals, Minnesota Twins</td>
</tr>
<tr>
<td>*West</td>
<td>Arizona Diamondbacks, Colorado Rockies, Los Angeles Dodgers, San Diego Padres, San Francisco Giants</td>
<td>Los Angeles Angels of Anaheim, Oakland Athletics, Seattle Mariners, Texas Rangers</td>
</tr>
</tbody>
</table>

Each team in the MLB plays 162 games in a season. The MLB season starts on the first Monday of April and runs through late September or early October. When the regular season ends, eight teams enter the post season. After elimination rounds the two remaining teams play a best best-of-seven World Series. The New York Yankees have been the most successful team in MLB history with 26 World Series Titles. The second closest is the St. Louis Cardinals with 10 Titles, followed by the Oakland Athletics with 9. Today, the MLB is led by stars such as Alex Rodriguez, Johan Santana, Derek Jeter, Barry Bonds, and Albert Pujols. These stars, along with some of the younger up and coming players, are going to continue to make this league one of the most popular and profitable sports in all of North America.
**Thesis Organization**

This thesis consists of five chapters. Chapter I is the introduction chapter. This chapter has provided the statement of purpose, the importance of the research, and a brief history of Major League Baseball. Chapter II is the literature review and model development chapter. This section will focus on previous research and studies that have been done in both the MLB and other major sports and also help define the model. Chapter III describes the research methodology. Chapter IV presents the results of data analysis. This chapter summarizes the results that were found and explains which variables effect MLB attendance. Chapter V summarizes the conclusions and describes the study’s contributions and offers suggestions for future implications.
CHAPTER II
LITERATURE REVIEW AND MODEL DEVELOPMENT

There have been numerous studies done on factors that affect attendance in most sports leagues. The four main professional sports leagues in America (Major League Baseball, National Football League, National Hockey League, and National Basketball Association) have been studied along with many collegiate and minor league sports. These studies set out to determine what factors affect attendance for a given sport. Even though there have been many studies throughout sports on attendance, each study uses different factors and some contain more updated data. The first section of this literature review will discuss studies that have analyzed sports attendance. This section will be divided up by different variables such as facility age and city population and the articles reviewed will have studied the specific variable. The second section will then make sense of all the studies and then a model for baseball attendance will be made.
Sports Attendance

Sporting events are considered part of the entertainment industry. In the entertainment industry there is competition for consumers’ leisure time and dollars, and over the last 20 years professional team sports have fared extremely well in the entertainment industry because attendance in sports has been on the rise. The NBA has increased its average per game attendance by almost 7,000 since 1980 and the MLB has increased its attendance by almost 10,000 in the same time period.¹

Winning Record

Winning games is the most important aspect of a professional sport. It is the reason most players play the game and the ultimate goal of all teams is to win a championship. Fans feel the same way. They want to see their team win every time they attend a game and it seems logical to think that the more a team wins the more fans it will have. Not surprisingly previous research has confirmed that the more a team wins the more likely they are to draw more fans. For example, Andrew Welki used a Tobit analysis to examine factors that affect game-day attendance in the NFL. Data for the study was collected for each game during the 1991 season. One of the variables examined was home team’s record. The Tobit analysis confirmed that by having a winning record, game-day attendance will be higher than teams that have losing records.²

In another study, D.A. Peel looks at the determinants of match attendances in the English Football (Soccer) League. Data for this study was collected from the English


Football League during the 1986-1987 season. Peel uses a least squares estimator and results show that home fans like to see their own team win in a high scoring game which is not too one sided. ³

A study done on the NBA also shows that fans prefer winning teams. David Berri examined three primary factors for the study that included team performance and star power. Berri used a regression equation and results show that star power is significant to consumer demand but fans still prefer to see their home team win more games.⁴

Iain Fyfee analyzed factors affecting attendance in the NHL. Fyfee specifically looked at goals scored, records, penalty minutes, location, and novelty. For this study a correlation test was used and the test found that a better record increased attendance.⁵

But does winning record play an important role in the MLB? The effects of a winning record in the MLB have not been studied in detail but will be a key variable in this study. Chapter I asked why the New York Yankees average almost 40,000 more fans than the Florida Marlins when both teams made the playoffs in 2005. Is this an unusual case or is it just that MLB fans don’t care whether or not their team wins? By looking at research done in other sports this thesis predicts that a winning record will have a positive effect on MLB attendance and the higher the wins a certain team has the greater the attendance should be.

---


Ticket Price

Fans attitudes usually depend a lot on ticket price. During almost every off-season, teams announce that their ticket prices are going up due to the increase in player salaries. The MLB has historically had the cheapest average tickets but prices continue to rise along with the rest of professional sport leagues. Because of this, the common theme throughout sports is that ticket price has a direct effect on attendance. If ticket prices were extremely high attendance might fall and vice versa. Previous research in the NFL has examined the effects of ticket prices.

In the NFL study by Andrew Welki, ticket price was one of the variables examined. Data for the study was collected for each game during the 1991 season and through a Tobit analysis, Welki found that higher ticket prices reduce attendance.

Craig Depkin studied fan loyalty in the NFL from 1990-1997. Depkin used attendance as his dependent variable and ticket price as one of his independent variables. Through a regression analysis, results showed that ticket price was negatively significant towards attendance.⁶

MLB has the cheapest tickets of the four major sports for a number of reasons. One reason the tickets are cheaper is because of the number of games played throughout a season. Each MLB team plays 81 home games, double the amount of NHL and NBA teams (NFL teams have 8 home games). MLB teams can afford to charge lower tickets because of the fact that they have so many home games. Another reason for the cheaper tickets is the fact that the average size of a stadium is 48,000 seats, which along with the number of home games, allows teams to have lower ticket prices. With that being said,

---

this thesis predicts that even though MLB ticket prices are lower than other sports, they will have a negative effect on attendance. Teams with higher ticket prices should have less attendance than those with cheaper tickets.

**Payroll**

The payroll of a team is something that a lot of fans pay attention to. Some feel the higher the payroll, the more wins a team will have. If a team has a higher payroll it means that more money is being spent on players. Because of this, fans like to see their teams spend millions of dollars and accumulate a high payroll.

"Star players, Payroll Distribution, and Major League Baseball Attendance," by Dominic Rivers looks at the impact star players have on attendance and the relationship between team payroll and attendance. Through a demand model Rivers concludes that a high payroll results in more fans but agrees that there are always teams with low payrolls that average a high number of fans. Rivers also concludes that if teams want to maximize their attendance, the team payroll should be distributed evenly across the roster instead of devoting a high percentage of the payroll to one or two players. 7

Because the MLB does not use a salary cap teams can spend as much or little on players as they want. During the 2007 season the New York Yankees had the highest payroll at $189,639,045 and the Florida Marlins had the lowest at $30,507,000. The wide range of payrolls makes this a perfect variable for the study and this thesis predicts that payroll will have a positive effect on attendance.

Facility Age

The last twenty years has seen a dramatic increase in the number of sport stadiums being built throughout North America. One of the reasons for building new stadiums is the hope that they will draw more fans. Sometimes part of attending a professional sport event is seeing the stadium and its new features.

Many studies have been done on facility age including one by Craig Depkin. Depkin’s study looked at fan loyalty and used stadium age as one of his independent variables. Depkin found that stadium age was negatively associated with attendance meaning the older the stadium the lower the attendance.\(^8\)

John Leadley studied the “honeymoon effect” on NBA teams with a new arena from 1971-2000. The “honeymoon effect” is essentially the idea that when a new stadium is built, there is a certain time period in which attendance will increase as a direct result of the new stadium. Using a cross section time-series sample, Leadley found that initial attendance increases 15% to 20% in the first 5 years of a new facility. The “Honeymoon effect” then begins to diminish after year five and has no affect by year ten.\(^9\)

According to several studies, the age of the stadium has a lot to do with drawing fans in the MLB. In an article written by Chad McEvoy, the relationship between facility age and attendance is examined. McEvoy also uses an economic demand model to examine the relationship between facility age and attendance, and he concludes that


attendance was highest during the first years of the new facility. Attendance was found to be lowest during the 48th year of the facility, followed by rising attendance as the facility continued to get older. In “Novelty Effects of New Facilities on attendance at Professional Sporting Events,” Dennis Coates shows that 2,500,794 additional tickets are sold over the first eight seasons of a new stadium. Coates uses a live game attendance model that concludes MLB attendance increases as a direct result of the construction of a new stadium or arena in a city.

In another study by Zenon Zygmont and John Leadley, the “honeymoon effect” is examined in Major League Baseball. Zygmont and Leadley’s study uses MLB data from the 1970 season to the 2000 season. This study uses a profit maximizing model that incorporates attendance and ticket price with several variables such as stadium age, team performance, and a dummy variable for location and year. The authors conclude that the “honeymoon effect” will last eight to ten years and a new baseball only park that replaces an older multipurpose stadium will generate an additional $228 million in ticket sales over a 15 year time period.

Previous research indicates that a new stadium does increase attendance but for only so long. The “honeymoon affect” seems to affect not only the MLB but other professional sports leagues as well. After looking at previous research this thesis predicts that a new stadium with have a positive effect on attendance but only for so long.

---


Ballparks that are 10 years and younger should show significantly higher attendance numbers than the teams with old stadiums.

**Geographical Location**

When a new team wants to join a professional league there are months of research that go into the process. Owners want to put teams in a city that would welcome and support a franchise. For a long time there were no NHL teams located in the south because hockey was not very popular in the certain region. Over the years, the NHL has expanded to the south but the attendance numbers are lower than teams located in the north where hockey is more popular. The location of a team is very important and previous research shows the effects it has on attendance.

Looking back, the study by Iain Fyffe analyzed factors that affect attendance in the NHL. Fyffe specifically looked at goals scored, records, penalty minutes, location, and novelty. Fyffe used a correlation test and results showed that location was a significant factor. Fyfee also concluded that a team’s attendance will be lower if a team is located in the south.\(^ {13}\)

A similar study was done by D.G. Ferguson and J. Jones. These two authors studied the effect of location, market power, and long term quality on attendance. The authors developed a model in order to clarify whether location and other factors explain a

team's profitability. Data was collected from the 1977-1978 season and after running the model the authors concluded that location had a significant effect on attendance.\textsuperscript{14}

MLB teams are located all over North America. It does not matter if a team is located in the south or north because baseball is played in the summer. Although some may argue that teams in the south have lower attendance during the intense summer months, geographical location of a baseball team is not that significant and will not be used as a variable. Similar variables such as number of substitute teams and city population will be discussed later in the chapter.

\textit{Substitute Teams}

New York City is host to seven main professional sport franchises, the most of any city in America. Los Angeles, San Francisco, and Baltimore come in second with six. Even though the cities are highly populated, one must wonder if all these substitute teams affect attendance.

Jason Winfree, uses a travel-cost model to analyze the attendance impacts on MLB compared to location. Winfree asserts that the closer two teams are to each other, the lower the attendance. Also, when a team moves closer to an existing team, there is an initial reduction in attendance to the current team. Winfree uses the Oakland Athletics as an example. In 1968, the Kansas City Athletics moved to Oakland causing San Francisco Giants attendance to drop by 32\%.\textsuperscript{15}


The more teams in one area should decrease attendance because fans have several sport options. Even though the MLB is played during the summer months there are still times when the seasons overlap. Therefore this thesis predicts that substitute teams will have a negative effect on attendance.

**Star Players**

Another reason fans attend sporting events is to root for their favorite star player. The more star players a team has, the more attractive that team would be. Having star players on a team, such as a Lebron James, makes the game more exciting for fans to watch. Before the arrival of King James, Cleveland drew fewer fans because of the fact that they did not have a star player. After James first season the attendance numbers dramatically increased even though the team won a similar amount of games.

David Berri examined the impact a star player has on attendance and used three primary factors for the study. These three factors include team performance, franchise characteristics, and market characteristics. Berri used a regression equation and results show that star power is significant to consumer demand but fans still prefer to see their home team win more games.\(^{16}\)

In December of 2000, Alex Rodriguez signed the largest player contract in the history of sports with the Texas Rangers. Rodriguez would receive $252 million over 10 years, including $22 million for the 2001 season. Texas Ranger owner Tom Hicks said, “We can pay this kind of money and make a profit.” (Rivers 2002; 165) Hicks’ staff then projected an attendance of 3 million fans, which would be the highest in MLB history.

---

The Rangers seemed to believe that a star player will increase attendance but what do MLB studies show?\(^{17}\)

"Star players, Payroll Distribution, and Major League Baseball Attendance," by Dominic Rivers looks at the impact star players have on attendance. Rivers developed a demand model that shows a star player only affects attendance if he helps contribute to the team’s success.\(^{18}\)

It seems obvious that more fans attend sporting events when there is a star player on the team. Previous studies support this claim and therefore this thesis predicts that the more star players a team has, the higher the attendance numbers will be. Star players should have a positive effect on attendance.

**Homeruns**

Baseball has become a power game since the emergence of steroids and HGH along with players working out more. This power game has resulted in more homeruns, something all fans enjoy seeing. The idea of catching a Barry Bonds homerun ball makes sitting in the outfield more enjoyable. Today, fans love seeing homeruns. Some enjoy watching the ball fly out of the park and others enjoy watching the players trot around the bases. But does more homeruns mean more fans? In the 2007 season the Milwaukee Brewers led the league with 231 homeruns and averaged 35,421 fans. The Kansas City Royals came in last with 102 homeruns and averaged 19,961. From this stat and the


popularity of homeruns, this thesis predicts that homeruns will have a positive effect on attendance.

**Strikes**

Over the past decade Major League baseball and the National Hockey League have been shut down because of player strikes and management lockouts. In 1994 the MLB cancelled the World Series because of a player strike and in 2006 the entire NHL season was lost due to a lockout. The problem that these leagues face is getting fans to return after the strike is over. Does losing an entire season due to a strike help or hurt attendance?

Several studies look at the affect of a labor strike on MLB attendance. “The Effects of Labor Strikes on Consumer Demand in Professional Sports,” by Victor Matheson concludes that the 1981 and 1994 MLB strikes caused short term attendance losses but have not resulted in any long term effects on attendance. Matheson wrote that even though attendance at MLB games following the 1994-1995 strike recovered to its pre-strike levels, it was result of new stadiums being built at a record pace. He then concludes that after taking new stadiums into consideration, MLB attendance has dropped significantly since the 1994-1995 strike.19

David Berri and Martin Schmidt study the impact of a strike on MLB attendance and whether or not a strike has permanent negative effects on attendance. For the study, total attendance for all 30 teams was taken from 1901-1998 and was scaled in terms of the 1992 attendance figures. Berri and Schmidt used a time-series analysis that studies the

---

1981 and 1994-1995 strikes. The results concluded that the strikes had only a small effect on attendance in the short run and there was no evidence to support the idea that strike causes permanent damage to MLB attendance.20

Dennis Coates and Thane Harrison studied the effects of lockouts on the demand for attendance by using data from 1969 to 1996. The authors used a regression model with several variables and found that the negative strike effects are significantly lower than the Berri and Schmidt study but still effect attendance negatively. The authors also found that the demand for attendance is price inelastic no matter how the basic ticket price variable is calculated.21

Promotions

Avid fans love the fact that they can attend a sporting event and in return take something home with them. Many professional sports teams use promotions to get more fans to attend games. In the MLB there are bobblehead and bat days, which means that a certain number of fans get one of these treasured items. Teams feel that if they offer the fans a gift for coming to the event, more spectators will show up. Previous research looks at the effects of promotions and whether or not they actually increase attendance.

Zhang, Pease, Hui, and Michaud examined several variables that are supposed to influence fans to attend an NBA game. For this study, the authors developed a Spectator Decision Making Inventory (SDMI) in order to test fans decision making. The SDMI included 20 variables that the authors believed would influence a fan, one being


promotions. Samples of 861 NBA fans were surveyed and a regression equation was used to analyze the data. Results show that promotions positively affected game day attendance. An article by Thomas Bovd studies promotions that increase the attractiveness of a baseball game. Bovd looks at three different types of promotions and their effects on attendance. Bovd combines the types of promotions with two timing factors, weekend games and games against rivals. Bovd concludes that promotions such as special features, giveaways, and price discounts have different effects on attendance. Bovd indicates that promotions don’t necessarily bring back fans or increase attendance by a significant amount.

In another study, Mark McDonald and Donald Rascher examine the effect that promotional days have on the attendance of a Major League Baseball game. The authors used an ordinary least squares regression and collected data from the 1996 MLB season for 19 teams. The results showed that special promotional days increased game day attendance by 14%. Also, by increasing the number of promotions a team has per year can sometimes have a negative effect on the impact of each promotion.

Playoff

Many professional sport teams consider a season a success if they make it to the playoffs. Even though it is every team’s goal to win a championship, only one team gets


to do that per year. Therefore teams become realistic and measure success by making the playoffs. But do fans feel the same way? If a team made the postseason the year before does it positively affect attendance?

Previous research is limited on the affects of playoffs on attendance. In the case of the MLB, only eight teams make the playoffs, making baseball one of the toughest sports to win a championship. Because previous research shows that fans enjoying winning, this thesis predicts that teams that playoffs will have a positive effect on attendance. Teams that made the playoffs the previous season should see higher attendance numbers than teams that did not make it.

City population

The thirty cities that host a MLB team range significantly in population size. The Pittsburgh Pirates play in the smallest city with a population of 313,278 and at 8,226,378, the New York Yankees and Mets play in the largest market. With such a drastic difference in population size it would be expected that the teams with higher populations should average more fans.

Craig Depkin studied fan loyalty in the NFL from 1990-1997. Depkin used attendance as his dependent variable and city population as one of his independent variables. Results showed that population was positively significant on attendance.²⁵

The MLB seems no different than the NFL in the wide range of city populations. This thesis predicts that population should have a positive influence on attendance throughout the MLB.

**Competitive Balance**

Competitive balance or outcome uncertainty is very important to the success of major league sports. Fans always love to see their home team win but they also like the fact that on any given day either team has a chance to win. Outcome uncertainty gives the fans a reason to come watch the game. Previous research looks at the effects of competitive balance on attendance.

David Forrest studied whether attendance in English soccer respond positively to the amount of uncertainty of the outcome between two teams in a match. Data were used from the 1999 season from four leagues with a total of 92 clubs. Results show that admissions at English soccer matches relate positively to the quality of teams playing each other and negatively to a measure of the win probabilities of each team. Also, although fans prefer an uncertainty of outcome, a greater quality of one team will have more affect on attendance than not knowing who will win.26

A similar study by D. A. Peel looks at the determinants of match attendances in the English Football (Soccer) League. Data for this study were collected from the English Football League during the 1986-1987 season. Peel uses a least squares estimator and results show that home fans like to see their own team win in a high scoring game which is not too one sided.27

Berri and Martin Schmidt examined the connection between aggregate attendance and MLB competitive balance. In order to determine the league’s competitive

---


balance, a Gini coefficient was used, along with the average league attendance from 1901 to 1998. Berri and Schmidt thought that the inconsistency in player salaries will reduce the competitive balance and therefore the league's attendance would be directly affected because it directly relates to competitive balance. By using a time series analysis, the authors conclude that competitive balance has a positive effect on league attendance and their hypothesis was correct.\textsuperscript{28}

\textit{Income}

The wealth of the population is a factor that could possibly affect attendance. If average income of a population is low in a city, fans might find it too expensive to attend a game. With the rise in ticket prices, fans really need to open up their wallets in order to bring a family of four to a game. The NBA, NFL, and NHL all have average ticket prices above $40.

Andrew Welki's study examined factors that affect game-day attendance in the NFL. Welki used data for each game during the 1991 season and examined income as one of his variables. Results reveal that income has little or no effect on attendance. Welki felt the reason for this was because the NFL regularly sold out most of its games, therefore income had little effect.\textsuperscript{29}

The average ticket price in the MLB is around $19, much lower than the NBA, NFL, and NHL. In the MLB income should have little to do with attendance because the


games are relatively affordable. This thesis predicts that income will have little (positive) or no effect on attendance.

**Interleague Play**

Interleague games always seem to have more meaning than a regular game. Interleague games give the fans a chance to see teams that they rarely see play. It would seem that attendance would increase during interleague play because of the chance to see a player like Barry Bonds or Alex Rodriguez in person. Michael Butler researched the difference between regular game day attendance and interleague play during the 1999 season. Butler used a regression equation that looked at the expected quality variables, current team performance variables, weather and time variables, and dummy variables. After running the regression, Butler concluded that there was a 7% increase in attendance during an interleague game.\(^\text{30}\)

**Review of Sport Studies**

After reviewing several studies done throughout sports, it is clear that many factors effect attendance at sporting events. What makes attendance so interesting is the fact that different sport fans go to games for different reasons. NHL fans prefer to attend NHL games for different reasons than an NFL fan. It is important to examine studies done throughout sports because it shows that different factors affect several sports. Table 2.1 shows the past research findings for the MLB, Football, Basketball, Soccer, and Hockey sporting events.

TABLE 2.1

Past Research on Factors Affecting Attendance in Sport Leagues

<table>
<thead>
<tr>
<th>Factor</th>
<th>MLB</th>
<th>Hockey</th>
<th>Football</th>
<th>Basketball</th>
<th>Soccer</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ticket Price</td>
<td>Not studied</td>
<td>Not studied</td>
<td>- (Welki Andrew, 1994)</td>
<td>Not studied</td>
<td>Not studied</td>
</tr>
<tr>
<td>Payroll</td>
<td>+ (Rivers Dominic, 2002)</td>
<td>Not studied</td>
<td>Not studied</td>
<td>Not studied</td>
<td>Not studied</td>
</tr>
<tr>
<td>Facility Age</td>
<td>+ (McEvoy Chad, 2005)</td>
<td>Not studied</td>
<td>- (Depkin Craig, 2001)</td>
<td>- (Leadley John, 2005)</td>
<td>Not studied</td>
</tr>
<tr>
<td>Geographical Location</td>
<td>Not studied</td>
<td>Mixed (Fyffe lain, 2002)</td>
<td>Not studied</td>
<td>Not studied</td>
<td>Not studied</td>
</tr>
<tr>
<td>Substitute Teams</td>
<td>- (Winfree Jason, 2004)</td>
<td>Not studied</td>
<td>Not studied</td>
<td>Not studied</td>
<td>Not studied</td>
</tr>
<tr>
<td>Homeruns</td>
<td>Not studied</td>
<td>Not studied</td>
<td>Not studied</td>
<td>Not studied</td>
<td>Not studied</td>
</tr>
<tr>
<td>Strikes</td>
<td>- (Matheson Victor, 2006)</td>
<td>Not studied</td>
<td>Not studied</td>
<td>Not studied</td>
<td>Not studied</td>
</tr>
<tr>
<td>Promotions</td>
<td>Mixed (Bovd Thomas, 2003)</td>
<td>Not studied</td>
<td>Not studied</td>
<td>+ (Zhang &amp; Pease, 1995)</td>
<td>Not studied</td>
</tr>
<tr>
<td>Playoff</td>
<td>Not studied</td>
<td>Not studied</td>
<td>Not studied</td>
<td>Not studied</td>
<td>Not studied</td>
</tr>
<tr>
<td>City population</td>
<td>Not studied</td>
<td>Not studied</td>
<td>+ (Depkin Craig, 2001)</td>
<td>Not studied</td>
<td>Not studied</td>
</tr>
<tr>
<td>Competitive Balance</td>
<td>+ (Berri &amp; Schmidt, 2001)</td>
<td>Not studied</td>
<td>Not studied</td>
<td>Not studied</td>
<td>+ (Forrest David, 2002)</td>
</tr>
<tr>
<td>Income</td>
<td>Not studied</td>
<td>Not studied</td>
<td>Mixed (Welki &amp; Ziapor, 1994)</td>
<td>Not studied</td>
<td>Not studied</td>
</tr>
<tr>
<td>Interleague Play</td>
<td>+ (Butler Michael, 2002)</td>
<td>Not studied</td>
<td>Not studied</td>
<td>Not studied</td>
<td>Not studied</td>
</tr>
</tbody>
</table>
Table 2.1 shows what has been studied for each sport and what has not. It is interesting to see that some factors affect two sports in different ways. Facility age has a positive effect on attendance in the MLB and NBA but has a negative effect in the NFL. Table 2.1 also shows that many variables have yet to be studied in certain sports.

**Review of Baseball Studies**

The review of previous research shows that there are indeed specific factors that affect attendance at a Major League baseball game, both in a positive and negative manner. Previous research found that *facility age, star players, promotions, payroll, and interleague play* had significant positive effects on attendance. Also, according previous research, *substitute teams, and strikes* were found to have negative effects on attendance. Even after looking at previous research there is still many factors that have yet to be examined. Important variables such as *winning record, ticket price, city population, city income, homeruns, and playoffs* have yet to be studied in depth.

**Summary**

This study will be a combination of past research on sporting events and new research with different variables old studies fail to mention. Past MLB studies fail to focus on several important variables that have been proven to effect attendance. This study is going to build on previous research by using the traditional variables such as *facility age, star players, and promotions* along with several new variables that will then give the best possible explanation of what factors effect MLB attendance.
Below is the model that is going to be used for this thesis. The dependent variable and the independent variables will be explained along with the expected outcome of each variable in Chapter III. The developed model that will be the basis for all tests done in this study:

$$\text{ATTENDANCE} = \beta_0 + \beta_1 \text{ YEAR} + \beta_2 \text{ ARENACAPACITY} + \beta_3 \text{ REGWINS} - \beta_4 \text{ AVG.TICKET} + \beta_5 \text{ PAYROLL} + \beta_6 \text{ FACILITY AGE} + \beta_7 \text{ SUBSTITUTIONS} + \beta_8 \text{ STAR PLAYERS} + \beta_9 \text{ TOTHOMERUNS} + \beta_{10} \text{ PLAYOFF} - \beta_{11} \text{ CITY POPULATION} + \beta_{12} \text{ INCOME}$$
CHAPTER III
METHODOLOGY

The previous chapter developed the model that is going to be the basis for all tests done in this study. The next step is to explain all the variables and the data that will be needed for the model. In doing so, the source of the data will be explained along with a time frame for the data collected.

**Dependent Variable**

Previous literature helped in determining the variables that are going to be used in the model. This model consists of a dependent variable and many independent variables. The function of the model is to figure out if the independent variables have an effect on the dependent variable. In the case of this research question the dependent variable is *Average Home Attendance*. Attendance numbers for all 30 MLB teams are easily accessible making the dependent variable very easy to find. For example, in 2006 the New York Yankees led the league in attendance by drawing an average of 52,739 fans. The Florida Marlins came in last with an average of 16,919 fans. Attendance numbers will be gathered from the Major League Baseball’s official website, MLB.com. For the purpose of this study, average attendance will be gathered from the 2003 MLB season to the 2007 season. This time frame allows five years of data to be studied, making the results more sufficient.

27
Control Variables

The model contains two control variables: Year and Arena Capacity. The Year variable is the year in which the data is collected. The first season studied was the 2003 season, then the 2004 and so on. The study collected data till the most recent season in 2007.

The Arena Capacity variable looks at the capacity of each team’s ballpark for that year. The capacities change over time due to the construction of new stadiums and the renovations of old ones. Ballpark capacity in the MLB range from as high as 62,623 (Vet Stadium) to a low of 34,007 (Oakland Stadium).

Independent Variables

For the model at hand, there will be ten independent variables used. The first independent variable is Regular Season Wins. It is easy to say that the winning teams average more fans than the losing teams but that might not always be the case. In 2005 the Chicago White Sox had the most regular season wins in the MLB, yet they didn’t even average 30,000 fans a game. During the same year the Los Angeles Dodgers had the fourth fewest wins but still managed to average 44,500 fans a game. MLB regular season records are available online at MLB.com and will be gathered from 2003 to the 2007 seasons.

The second independent variable is Ticket Price. For this variable, every team’s ticket prices will be examined at. Most ballparks have a range of ticket prices. Usually the premium seats are the most expensive and “nose bleeds” are the cheapest. The
average ticket price for each team will become the number that is used. Data will be compiled from teammarketing.com and data will be from the last five seasons.

The third independent variable is **Team Payroll**. Every team’s payroll for the last five years will be gathered. MLB payrolls range from approximately $180,000,000 to $24,000,000. Team payroll data will be found at USAToday website.

**Facility Age** is the fourth independent variable. Team facilities in the MLB range from new state of the art stadiums to old ballparks with tons of character and history. The Yankees, Cubs, and Red Sox all play in old ballparks with character, yet still average a lot of fans. At the same time, teams with newer ballparks average a lot of fans because fans enjoy watching a game in a brand new stadium. Ballparksofbaseball.com gives detailed descriptions of every team ballpark. The opening and closing dates of ballparks are given, along with capacity size.

**Substitute Teams in an Area** is another variable. The number of teams within a 60 mile range is considered to be a substitute. Professional leagues that are considered to be substitutes are the National Hockey League, the National Football League, and the National Basketball Association. Data for substitute teams per city will be found at Wikipedia online. The New York Yankees and Mets have the most substitute teams at seven whereas the Milwaukee Brewers have the least with one.

The sixth independent variable is **Star Players**. This variable looks at the impact a star players has an attendance. A player is considered a star player if he made it to the previous all-star game. The all-star rosters for each season provide this data. This data will be compiled from the MLB’s official website. League rules require at least one player per team to make the all-star game every season, so there will never be a team
without an all-star. In the 2004 season, the New York Yankees had eight all-star players, the most in the past five seasons.

*Total Home Runs* is the number of home runs a team accumulated over a season. Team’s home run totals will vary from year to year.

Whether or not the team made the *playoffs* the previous season is another variable. Team interest usually intensifies when teams have the opportunity to win a championship. The playoff variable will show whether playoff excitement carries over to the next season. This variable is not a numerical factor therefore it requires the use of a dummy variable. A value of 1 is given to all teams that make playoffs and a value of 0 is given to the rest. Dummy variables allow non numerical factors to be studied in the given model.

The data for *City Population* and *Medium Household Income* are found at the United States Census Bureau for each city from the last five years.

Table 3.1 on the following page displays the variables that will be studied in this thesis along with their expected outcomes. Table 3.1 also lists the data source and time frame for which the data will be gathered.
**TABLE 3.1**

<table>
<thead>
<tr>
<th><strong>Variable</strong></th>
<th><strong>Expected Outcome</strong></th>
<th><strong>Data Source/Time frame</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>Regular Season Wins</td>
<td>Positive</td>
<td>MLB.com/ 5 Years</td>
</tr>
<tr>
<td>Ticket Price</td>
<td>Negative</td>
<td>Teammarketing.com/ 5 Years</td>
</tr>
<tr>
<td>Team Payroll</td>
<td>Positive</td>
<td>USAtoday.com/ 5 Years</td>
</tr>
<tr>
<td>Facility Age</td>
<td>Positive</td>
<td>Ballparksofbaseball.com/ 5 Years</td>
</tr>
<tr>
<td>Substitute Teams</td>
<td>Negative</td>
<td>Wikipedia/ 5 Years</td>
</tr>
<tr>
<td>Star Players</td>
<td>Positive</td>
<td>MLB.com/ 5 Years</td>
</tr>
<tr>
<td>Total Home Runs</td>
<td>Positive</td>
<td>MLB.com/ 5 Years</td>
</tr>
<tr>
<td>Playoff</td>
<td>Positive</td>
<td>MLB.com/ 5 Years</td>
</tr>
<tr>
<td>City Population</td>
<td>Positive</td>
<td>US Census Bureau/ 5 Years</td>
</tr>
<tr>
<td>Medium Household Income</td>
<td>Positive</td>
<td>US Census Bureau/ 5 Years</td>
</tr>
</tbody>
</table>

**Summary**

This chapter explained all the variables that are going to be used in this study and in doing so, the source of the data and the time frame was presented. The next chapter will report the results from the models along with explanations of which variables affect attendance.
CHAPTER IV
RESULTS

Table 4.1 displays the Summary Statistics. This table lists the dependent variable and the 12 independent variables. From the data gathered, the mean, standard deviation, minimum value, and maximum value are presented. This data was gathered from 150 observations.

TABLE 4.1
Summary Statistics

<table>
<thead>
<tr>
<th>Variable</th>
<th>Mean</th>
<th>Standard Deviation</th>
<th>Min</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>Attendance</td>
<td>30,699.11</td>
<td>8654.64</td>
<td>9,356</td>
<td>52,739</td>
</tr>
<tr>
<td>Year</td>
<td>2005</td>
<td>1.418</td>
<td>2003</td>
<td>2007</td>
</tr>
<tr>
<td>Arena Capacity</td>
<td>45,469.69</td>
<td>6277.13</td>
<td>34,007</td>
<td>62,623</td>
</tr>
<tr>
<td>Regular Season Wins</td>
<td>80.98</td>
<td>11.39</td>
<td>43</td>
<td>105</td>
</tr>
<tr>
<td>Avg. Ticket Price</td>
<td>21</td>
<td>6.33</td>
<td>10</td>
<td>49</td>
</tr>
<tr>
<td>Payroll</td>
<td>74,646,757.61</td>
<td>32277584.55</td>
<td>14,998,500</td>
<td>208,306,817</td>
</tr>
<tr>
<td>Facility Age</td>
<td>23.32</td>
<td>25.65</td>
<td>0</td>
<td>95</td>
</tr>
<tr>
<td>Substitute Teams</td>
<td>3.17</td>
<td>1.587</td>
<td>1</td>
<td>7</td>
</tr>
<tr>
<td>Star Players</td>
<td>2.24</td>
<td>1.48</td>
<td>1</td>
<td>8</td>
</tr>
<tr>
<td>Total Homeruns</td>
<td>173.45</td>
<td>32.30</td>
<td>102</td>
<td>260</td>
</tr>
<tr>
<td>Playoff</td>
<td>.27</td>
<td>.436</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>City Population</td>
<td>1,515,888.33</td>
<td>2004917.93</td>
<td>312,819</td>
<td>8,226,378</td>
</tr>
<tr>
<td>Income</td>
<td>40,771.81</td>
<td>8470.12</td>
<td>22,978</td>
<td>65,497</td>
</tr>
</tbody>
</table>

Table 4.2 on the following page displays the Correlation Matrix from the variables used in the five models.
<table>
<thead>
<tr>
<th>Table 4.2: Correlation Matrix</th>
</tr>
</thead>
<tbody>
<tr>
<td>Average Attendance</td>
</tr>
<tr>
<td>Arena Capacity</td>
</tr>
<tr>
<td>Avg. Ticket Price</td>
</tr>
<tr>
<td>Facility Age</td>
</tr>
<tr>
<td>Income</td>
</tr>
<tr>
<td>Payroll</td>
</tr>
<tr>
<td>Players</td>
</tr>
<tr>
<td>Playoff</td>
</tr>
<tr>
<td>City Population</td>
</tr>
<tr>
<td>Reg. Season Wins</td>
</tr>
<tr>
<td>Substitute Teams</td>
</tr>
<tr>
<td>Total Homeruns</td>
</tr>
<tr>
<td>Year</td>
</tr>
</tbody>
</table>

33
Correlations among the dependent and independent variables are shown in Table 2. The first column shows what independent variables have high correlations with the dependent. A strong correlation among the two would be considered greater than .2. All but two of the independent variables show strong correlation with average attendance. The highest correlation is between payroll and average attendance at .79. This noticeably high correlation says that payroll and average attendance almost directly affect each other. This high correlation leads to the question of what variable causes the other? Is average attendance high because of payroll or is payroll high because of average attendance? (These questions, among others, will be addressed in the conclusions chapter.)

Table 2 also looks at the relationship between two independent variables. When two independent variables have a coefficient greater than .5, it means the variables are almost measuring the same thing, which can become a problem when running the regression. Table 4.2 has several values greater than .5 which means the model has a multicolinearity problem. Multicolinearity leads to problems when running the regression such as large variances and small t-stats. When multicolinearity occurs the best way to solve it is by combining one or more variables. Fortunately, even though this model has multicolinearity, the regression results were still significant at a high level, allowing the correlation matrix to stay the way it is.

The chapter now presents the results of the regression analyses. These analyses are based on the models developed at the end of Chapter II. This thesis ran five different
sets of models and each will be described and analyzed throughout this chapter. The first model examined was the effect of the two control variables on attendance.

MODEL 4.1
ATTENDANCE = β0 + β1 YEAR + β2 ARENACAPACITY

The second model examined all the independent variables on attendance. This is the same model as described at the end of Chapter II and has been the base for all tests done in this study.

MODEL 4.2
ATTENDANCE = β0 + β1 YEAR + β2 ARENACAPACITY + β3 REGWINS -
β4 AVG.TICKET + β5 PAYROLL + β6 FACILITY AGE + β7 SUBSTITUTIONS +
β8 STAR PLAYERS + β9 TOTHOMERUNS + β10 PLAYOFF - β11 CITY
POPULATION + β12 INCOME

The third model studied uses only the variables that were significant from Model 4.2. Variables were considered significant if their t-statistic value was higher than the T-critical value of 1.96.

MODEL 4.3
ATTENDANCE = β0 + β1 AVG.TICKET + β2 FACILITY AGE + β3 PAYROLL+ β4 REGWINS + β5 SUBSTITUTIONS
The fourth model examined all the independent variables except payroll. The payroll variable was left out because it was considered the most significant when studied in Model 4.2.

MODEL 4.4
\[ \text{ATTENDANCE} = \beta_0 + \beta_1 \text{YEAR} + \beta_2 \text{ARENACAPACITY} + \beta_3 \text{REGWINS} - \beta_4 \text{AVG.TICKET} + \beta_5 \text{FACILITY AGE} + \beta_6 \text{SUBSTITUTIONS} + \beta_7 \text{STAR PLAYERS} + \beta_8 \text{TOTHOMERUNS} + \beta_9 \text{PLAYOFF} - \beta_{10} \text{CITY POPULATION} + \beta_{11} \text{INCOME} \]

The fifth model studied uses only the variables that were significant from Model 4.4. Variables were considered significant if their t-statistic value was higher than the T-critical value of 1.96.

MODEL 4.5
\[ \text{ATTENDANCE} = \beta_0 + \beta_1 \text{ARENACAPACITY} + \beta_2 \text{AVG.TICKET} + \beta_2 \text{FACILITY AGE} + \beta_3 \text{INCOME} + \beta_4 \text{REGWINS} + \beta_5 \text{SUBSTITUTIONS} \]

Table 4.3 on the next page displays the results from the first three regression analyses completed.
TABLE 4.3

Regression Analysis Results
(Independent variable in all models is Average Home Attendance, 
t-statistics in parentheses)

<table>
<thead>
<tr>
<th>Variables</th>
<th>Model #1</th>
<th>Model #2</th>
<th>Model #3</th>
</tr>
</thead>
<tbody>
<tr>
<td>C (Constant)</td>
<td>-2316581**</td>
<td>-397068 (-0.66)</td>
<td>-28.73 (-0.008)</td>
</tr>
<tr>
<td>Year</td>
<td>1158.78**</td>
<td>194.76 (0.64)</td>
<td></td>
</tr>
<tr>
<td>Arena Capacity</td>
<td>0.52***</td>
<td>0.14 (1.67)</td>
<td></td>
</tr>
<tr>
<td>Regular Season Wins</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Avg. Ticket Price</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Payroll</td>
<td>0.0001**</td>
<td>0.0001**</td>
<td></td>
</tr>
<tr>
<td>Facility Age</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Substitute Teams</td>
<td>840.33**</td>
<td>969.81**</td>
<td></td>
</tr>
<tr>
<td>Star Players</td>
<td>302.96 (0.85)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total Homeruns</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Playoff</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>City Population</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Income</td>
<td>0.06 (1.26)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>R-Squared</td>
<td>0.17</td>
<td>0.70</td>
<td>0.68</td>
</tr>
<tr>
<td>Adjusted R-Squared</td>
<td>0.16</td>
<td>0.68</td>
<td>0.67</td>
</tr>
<tr>
<td>F-Statistic</td>
<td>15.60***</td>
<td>27.90***</td>
<td>62.47***</td>
</tr>
<tr>
<td>N</td>
<td>150</td>
<td>150</td>
<td>150</td>
</tr>
<tr>
<td>P &lt; .05 *</td>
<td>P &lt; .01 **</td>
<td>P &lt; .001 ***</td>
<td></td>
</tr>
</tbody>
</table>
Model 1:

The first model contained only two independent variables, year and arena capacity. Both variables were statistically significant at the 5% level or better. The year variable had a t-statistic value of 2.53 with a coefficient of 1158.78. This statistic implies that for each year after 2003 just being in the league would increase a team’s attendance figures.

The other significant variable was arena capacity, which produced a t-statistic of 5.08 and a coefficient of 0.52. What this means is that arena capacity is very significant according to Model 1. The statistic implies that if a ballpark increases its capacity by 100 seats, the attendance will increase by 52 people holding all other variables constant.

Even though Model 1 produced two significant variables at the T-critical value level, it created an F-statistic of 15.60 and an adjusted r-squared of 0.16, the lowest of the five models examined. In all, this model explains 16% of the variation in average home attendance at MLB games during a five year span.

Model 2:

The second model examined all 10 independent variables and both control variables for a total of 12 variables. Of the 12 variables examined, five produced significant t-statistics. These significant variables are regular season wins, average ticket price, payroll, facility age, and substitute teams.

As predicted in Chapter II, regular season wins was one of the variables that would have a positive effect on attendance. This variable in Model 2 produced a t-statistic of 3.05 with a coefficient of 175.08. From the five significant variables, regular
*season wins* is the second most important variable that affects attendance. These statistics imply that one more regular season win will attract 175 more spectators.

*Average ticket price* proved to be a significant variable but in a positive direction. Previous chapters predicted that the average ticket price would negatively affect attendance because it seems logical to think the higher the ticket price the fewer fans a team will attract and vice versa. Model 2 produced an unexpected statistic that gives the *average ticket price* variable a t-stat of 2.96 and a coefficient of 300.39. This statistic implies that for a $1 increase in average ticket price, average home attendance will increase by 300. By looking at the 2007 MLB season, six teams averaged more than 40,000 fans. Of those six teams, four had average ticket prices above $25. What this means is teams that draw the most fans usually have the highest ticket prices. So when looked at this way, it seems reasonable to think that *average ticket price* could positively affect attendance.

Chapter II predicted that *payroll* would positively affect attendance and according to Model 2 it defiantly does. *Payroll* proved to be the most important variable that affects *average home attendance* with a t-statistic of 6.13 and a coefficient of 0.0001. The produced t-statistic of 6.13 is by far the greatest value of all 12 variables. It seems that when a team’s *payroll* increases by $10,000 average attendance increases by one spectator. Being that team payrolls are almost as high a $200,000,000 it seems reasonable to think that *payroll* could have such an effect on attendance.

According to Model 2 another significant variable that affects MLB attendance is *facility age*. This variable produced a t-stat of -2.64 with a coefficient of -59.50, the first significant variable that negatively affects attendance. Given the negative significant t-
stat it appears that the older a facility, the less fans. The statistics imply that for every year a facility gets older, average attendance will decrease by 59 fans. In other words, a ten year older ballpark will average 590 less fans than it did when the facility opened.

The last significant variable in Model 2 is the number of substitute teams. Previous chapters predicted that substitute teams would have a negative effect on attendance but according to Model 2 it actually is positively significant. This factor produced a t-statistic of 2.13 and a coefficient of 840.33. Out of all the significant variables, substitute teams has the lowest t-stat and closest to the t-critical value of 1.96 but is still considered significant. These statistics suggest that for an increase in one substitute team, average attendance would increase by 840. When looked at further, it seems that more substitute teams makes for a more sport oriented city. According to the model it seems that the more substitute teams in an area the greater attendance will be.

The remaining seven variables examined were insignificant, yet some produced values that were not expected. Total homeruns produced a t-statistic of -1.82, not quite enough to be considered significant, but surprisingly produced a negative number. The coefficient of -27.16 suggests that one more total homerun would decrease attendance by 27 fans. Total homeruns was predicted to be not only significant but positive. The other factor that went against initial predictions was the Playoff variable. Even though this variable was considered insignificant, it produced a surprising coefficient of -796.05. What this suggests is if teams made the playoffs the previous season attendance will automatically decrease by 796. Logically, one would think that a good playoff run the previous season would lead to excitement for the next year causing attendance to increase. One possible reason for this unexpected outcome is the fact that the playoff
variable was used as a dummy variable and given a value a 0 or 1. The results could be different if the number of playoff wins were accounted for instead of a yes or no approach.

Model 2 as a whole produced an F-statistic of 27.90, much higher than Model 1, and an adjusted r-squared of 0.68. The adjusted r-squared of 0.68 implies that the model explains 68% of the variation in average home attendance, much higher that the adjusted r-squared of 0.16 that Model 1 produced.

Model 3:

The third model took all the variables from Model 2 that produced significant t-statistic and left out the variables that were statistically insignificant. There were five variables from Model 2 that produced significant t-stats. Of the five variables, all remained significant, and four of them increased their t-statistic. The four variables that increased their t-stat were regular season wins, payroll, facility age, and substitute teams.

In Model 3, regular season wins produced a t-statistic of 3.65 and a coefficient of 148.57. The new results produced a higher t-stat but smaller coefficient. Instead of one more win attracting 175 more fans, Model 3 predicts only an increase of 148.

Payroll increased its t-statistic the most going from 6.13 to 7.17 while the coefficient remained the same at 0.0001. Model 3 reconfirms the significance of the Playoff variable and its enormous affect on average home attendance. The third variable that increased its t-stat is facility age. The new model produced a t-statistic of -2.82 and a coefficient of -62.46. The coefficient increased from Model 2 and now for every year the facility gets older, average attendance decreases by 62 fans.
The last variable is substitute teams. The new model produced a t-statistic of 2.69, up from Model 2, and the coefficient also increased from 840.33 to 969.81. The new model predicts that for an increase in one substitute team, attendance will increase by 969 fans.

The only variable that did not increase its t-statistic was average ticket price. The new model produced a t-stat of 2.69, down from 2.96, and a coefficient of 255.35. This variable is still considered significant but both its t-stat and coefficient declined.

In all, this model produced an F-statistic of 62.47, the most of the three models, and an adjusted r-squared of 0.67. The new adjusted r-squared remained consistent with Model 2 implying that the new model explains 67% of the variation in MLB attendance over a five year period.

Table 4.4 on the next page displays the results from the last two regression analysis completed followed by a brief description of each model.
TABLE 4.4

Regression Analysis Results
(Independent variable in all models is Average Home Attendance,
t-statistics in parentheses)

<table>
<thead>
<tr>
<th>Variables</th>
<th>Model #4</th>
<th>Model #5</th>
</tr>
</thead>
<tbody>
<tr>
<td>C (Constant)</td>
<td>-728780.8</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(-1.08)</td>
<td></td>
</tr>
<tr>
<td>Year</td>
<td>354.41</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(1.05)</td>
<td></td>
</tr>
<tr>
<td>Arena Capacity</td>
<td>0.25**</td>
<td>0.28**</td>
</tr>
<tr>
<td></td>
<td>(2.81)</td>
<td>(3.43)</td>
</tr>
<tr>
<td>Regular Season Wins</td>
<td>209.30**</td>
<td>223.62***</td>
</tr>
<tr>
<td></td>
<td>(3.26)</td>
<td>(5.24)</td>
</tr>
<tr>
<td>Avg. Ticket Price</td>
<td>595.17***</td>
<td>637.24***</td>
</tr>
<tr>
<td></td>
<td>(5.93)</td>
<td>(7.00)</td>
</tr>
<tr>
<td>Facility Age</td>
<td>-55.34**</td>
<td>-55.85**</td>
</tr>
<tr>
<td></td>
<td>(-2.19)</td>
<td>(-2.30)</td>
</tr>
<tr>
<td>Substitute Teams</td>
<td>1732.15***</td>
<td>1858.98***</td>
</tr>
<tr>
<td></td>
<td>(4.20)</td>
<td>(5.13)</td>
</tr>
<tr>
<td>Star Players</td>
<td>497.27</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(1.25)</td>
<td></td>
</tr>
<tr>
<td>Total Homeruns</td>
<td>-14.28</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(-0.86)</td>
<td></td>
</tr>
<tr>
<td>Playoff</td>
<td>25.23</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.017)</td>
<td></td>
</tr>
<tr>
<td>City Population</td>
<td>0.0001</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.48)</td>
<td></td>
</tr>
<tr>
<td>Income</td>
<td>0.11*</td>
<td>0.11**</td>
</tr>
<tr>
<td></td>
<td>(1.96)</td>
<td>(2.15)</td>
</tr>
<tr>
<td>R-Squared</td>
<td>0.62</td>
<td>0.62</td>
</tr>
<tr>
<td>Adjusted R-Squared</td>
<td>0.60</td>
<td>0.60</td>
</tr>
<tr>
<td>F-Statistic</td>
<td>21.34***</td>
<td>38.98***</td>
</tr>
<tr>
<td>N</td>
<td>150</td>
<td>150</td>
</tr>
<tr>
<td>P &lt; .05 *</td>
<td>P &lt; .01 **</td>
<td>P &lt; .001 ***</td>
</tr>
</tbody>
</table>
Model 4:

In the first three models payroll was the most significant variable among the 12 that were studied. The t-statistic was well above any other significant t-stat and the amount of correlation with other variables was highest. Model 4 simply removed the payroll variable and examined the effects of the rest of the variables on attendance.

Of the 11 variables examined, six produced significant t-statistics. With the payroll variable gone, both arena capacity and income became significant when compared to Model 2. The four other significant variables were average ticket, facility age, regular season wins, and substitute teams. With the absence of payroll, average ticket produced the highest t-statistic at 5.93 with a coefficient of 595.17. This statistic implies that for every $1 increase in ticket price, average home attendance will increase by 595 fans. Even though this logic contradicts common sense, it reconfirms, that according to this study, average ticket price positively affect attendance.

In total, Model 4 produced an F-statistic of 21.34 and an adjusted r-squared of 0.62, significantly lower than Model 2. This model explains 62% of the variation in MLB average home attendance.

Model 5:

Model 5 used only the six significant variables from Model 4. These variables include average ticket, arena capacity, income, facility age, regular season wins, and substitute teams. Each variable increased their t-statistic and all remained significant. In Model 4 regular season wins had a t-statistic of 3.26 and a coefficient of 209.30. Model 5 produced a t-statistic of 5.24 and a coefficient of 223.62, implying that for one more
win, average attendance will increase by 223 fans. Even though *regular season wins* showed the biggest t-statistic increase from the two models, *average ticket price* remained the most significant with a t-statistic of 7.00 up from 5.93.

Model 5 produced an F-statistic of 38.98 and an adjusted r-squared of 0.62. When compared to Model 4, the adjusted r-squared is the exact same but the F-statistic differ. The consistent adjusted r-squared implies that like Model 4, this model explains 62% of the variation in MLB attendance.

**Summary**

This chapter provided the results from the five models used. After running the regression and analyzing the results it became clear that there are certain variables that affect Major League Baseball attendance. The next chapter discusses the meaning of the results and further research that can be done in the future while concluding this thesis.
CHAPTER V
CONCLUSION

This final chapter will discuss the implications of the results along with the limitations and future research. The purpose of this thesis was to study the determinants of Major League Baseball attendance during a span of five years. Twelve variables were examined that could possibly affect MLB attendance and five regression models were used to test these variables. The twelve variables that were tested include year, arena capacity, regular season wins, average ticket price, payroll, facility age, substitute teams, star players, total homeruns, playoff, city population, and income. After running many regression analyses, it was concluded that seven variables significantly affect Major League Baseball attendance.

Summary

Arena capacity plays a role in determining a team’s average home attendance. Teams that have large ballparks are able to fit more seats giving more fans the option of coming to a game. The Boston Red Sox play in a ballpark that holds 36,000 fans, the lowest of any team in the MLB. Even though the Red Sox sell out every game, average home attendance numbers are much lower than most of the teams. If the Red Sox were to build a larger stadium, their average attendance would increase because of the strong demand for tickets in the Boston area.
Previous studies have always concluded that a winning team is one of the strongest factors in increasing attendance. This study was no different, as regular season wins was the second most significant variable throughout the study. The more regular season wins a team has the better their attendance will be. One would also think that a playoff team the previous season would positively affect attendance, but that was not the case in this study. The playoff variable was considered insignificant in all five models and it seems that the previous season does not really matter. This could account for the many “fair weather fans” that teams accumulate throughout a season. A “fair weather fan” would be someone who only likes the team when they are winning. This trend tends to happen late in the season when it looks like a team has a chance to win.

The most surprising significant variable was average ticket price. Historically it has been believed that the higher the ticket price, the less amount of fans. This study proved otherwise, having average ticket price positively affect attendance in four of the five models.

The most influential variable of the 12 that were tested turned out to be the payroll of a franchise. This variable was so powerful that it was removed from two of the models to see what the results would be without its presence. It makes sense for payroll to strongly affect attendance because the higher the payroll, the better the players will be. When the players are better it usually results in more wins, causing attendance to increase. The payroll variable also had the highest correlation with average home attendance when put into a correlation matrix, proving the two variables directly relate to one another.
Facility age was the only variable that was found to negatively affect attendance. It seems that fans are attracted to newer ballparks that have better technology and luxury. Older ballparks tend to be more uncomfortable for a fan, which could possibly push away some fans from attending a game. Major League Baseball owners are beginning to understand this as the last decade has seen the most new ballparks of any era in the history of the game.

It seems that sports specific cities attract more MLB fans. The more professional sport teams a city has, the higher the attendance will be according to this study. Substitute teams were found to be significant and some could argue that the result were a bit surprising. Logically, it would seem that if a city only had one sport franchise, attendance would be through the roof. But according to the results it seems that the more franchises a city have, the more sport fans in general, resulting in higher attendance.

The last significant variable was income. Even though this variable was found to be significant in one model it is still worth noting that income does positively affect attendance. Cities with higher medium household incomes are more likely to attract fans. The reason this variable was not significant in all the models is probably because baseball is a relatively cheap sport to watch, which make it affordable for almost anyone to go.

Seven of the twelve variables studied were concluded to be significant in at least one of the models examined. Of the seven, payroll and regular season wins proved to be the most affective. Even though this study produced great results there were still many limitations that this study faced.
Limitations of the Study

This study has proven to be an effective tool in predicting the determinants that affect MLB attendance. Even though the results are considered a success, there were still many limitations to this thesis. One limitation was time. If there was more time to complete this thesis, there could have been more observations gathered. Data for this study were gathered for the five year period, 2003 to 2007. Five years proved to be adequate for the purpose of this study but more years of data could show if there are longer term trends that occur in the MLB. Also if presented with more time, different variables could be looked at, some of which were left out of this study.

The income variable also posed a problem. Initially all of the medium household incomes were found at the United States Census Bureau. Data for the year 2007 was not yet available at the time of this study and in order to properly examine the income variable, the 2007 figures would be needed.

Implications for Theory

Previous research has found both similarities and differences with the results found in this study. In Andrew Welki’s study, it was found that ticket price negatively affects game day attendance and other studies have concluded similar results. Sport economists will be surprised to see these finding. In every model used, ticket price was found to be positively significant. Even when the data was changed in ordered to account for the “surprising” result, ticket price remained unchanged.

Another result that some economists might find surprising is the significance of the payroll variable. In the two models the variable was included, it showed the highest t-statistic of all the results. The payroll variable also showed the highest correlations with a number of other variables, including the dependent variable. One possible reason for the surprising results from both ticket price and payroll is the concept of causality. What this concept implies is that there is logical relationship between one variable to another variable in terms of a cause and effect connection. Throughout this thesis it has been presumed that the dependent variable is a direct result of all the independent variables. For example, attendance is high because the team has a high payroll. But is it the other way around? Is payroll high because the team attracts a lot of fans? Also, are high ticket prices a result from having higher attendance? These questions are valuable questions that could implicate different meaning to the results found. Some could argue that these results have minimal impact in the world of baseball because of the concept of causality.

Implications for Owners

Major League Baseball attendance has increased every season since 2003. In last year’s season, the MLB set a record with a total of 77,532,108 fans. This figure was up over 4 million from the previous season, so it seems as if baseball attendance is headed in the right direction. Although attendance is high, there are still many teams that struggle to attract fans. This thesis can provide valuable information to MLB team owners and management. Over the course of this these, different variables have been discussed in length and results showed what variables affect attendance. Owners sometime use common sense rather than using results. This thesis provides some surprising data that if
used properly, could benefit many teams. The reason this study proved to be so significant is because there are certain variables that affect attendance. Spectators attend games for certain reason and these reasons have been outlined in this study.

Major League Baseball is headed in the right direction. Baseball franchises are one of the most profitable franchises in all of sports and attendance is the number one reason for that. Even though the future looks bright, there is always room for improvement.

**Future Research**

This thesis is a great starting point for future research to build upon. This study uses twelve variables but many more could be added. Important variables such as game-day promotions, weather, number of double plays, and championships won could be added to improve the results. Also, the age of the franchise could be examined. Future research could also study the last ten years instead of five. Other sport attendance studies could also use these results in future studies regardless of what sport is being studied.

The purpose of this study was to identify the determinants that affect Major League Baseball attendance. This study built upon previous research done throughout sports in order to get the most updated and significant results. The information presented in this thesis could be useful to the owners and managers of the MLB because baseball is essentially a business. This thesis has identified several factors that affect attendance, and with this data, individuals such as owners and managers can focus their attention on certain aspects of the game in order to keep baseball “America’s Pastime”
SOURCES CONSULTED

JOURNAL ARTICLES


**WEBSITES**

Available at: http://www.ballparksofbaseball.com/attendance

Available at: http://www.census.gov

Available at: http://www.canada.com

Available at: http://www.mlb.com

Available at: http://www.teammarketing.com
Available at: http://www.usatoday.com

Available at: http://wikipedia.com