IF YOU WIN, THEY WILL COME: MAJOR LEAGUE BASEBALL AND THE UNCERTAINTY OF OUTCOME HYPOSTHESIS

by

Christopher Aemil Pouler

A thesis submitted to the faculty of
The University of North Carolina at Charlotte
in partial fulfillment of the requirements
for the degree of Master of Science in
Economics

Charlotte

2015

Approved by:

____
Dr. Craig A. Depken II

____
Dr. Louis Hawkins Amato

____
Dr. Arthur L Zillante

ABSTRACT

CHRISTOPHER AEMIL POULER. If you win, they will come: major league baseball and the uncertainty of outcome hypothesis. (Under the direction of DR. CRAIG A. DEPKEN II)

This paper extends the Knowles et al. (1992) paper titled “The Demand for Major League Baseball: A Test of the Uncertainty of Outcome Hypothesis.” The main research question is to replicate, and improve, the original paper using data from the 2013 MLB season. The main improvement from the original paper is using money lines to create a subjective probability of the home team winning. The logic behind using the money lines and odds is that consumers are more likely to attend a game when the home team has a significant chance of winning. The replication examines if the relationship between attendance and the probability of the home team winning still exists after 25 years and substantial changes to the league. This paper improves upon the original methodology to include different modeling techniques, including the use of panel data to control for team and time variation, and introducing different independent variables such as interleague play, games against rivals, the number of wins each pitcher has, the betting over/under line, and the current win/loss steak of each team. These changes investigate whether any of these additional variables or new modeling techniques show that subjective winning probability is still relevant to maximizing attendance.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER ONE: BACKGROUND

Sport leagues have a large impact on our lives. At least one sport has a game being played at any given time during the year; whether it is simply something on TV in the reception room of your doctor's office, or a nationally televised spectacle of a world championship game with multi-million dollar 30 second television ads. Casual and hardcore fans of each sport spend varying amounts of time and money to watch any or all the games that they can. Whether that means attending the games in person or paying for specific sports packages from their TV provider, this large demand has created a multibillion dollar industry. According to a study by Forbes, Major League Baseball saw a revenue figure of approximately $8 billion dollars for the 2013 season. This included naming rights of stadiums, ticket revenue, and lucrative television deals with TBS, FOX, and ESPN.

How can Major League Baseball maintain consumer interest in its product so to maintain their large revenue? This profit maximization strategy depends solely on the demand of consumers to watch the games. If consumers stopped attending games or watching games on TV, there would be no naming rights, ticket revenue, or TV deals. The main purpose of this paper is to provide an analysis of Major League Baseball and determine if home field advantage contributes to maximizing attendance for their games. This paper will replicate the findings from Knowles et al. (1992), and then modify their model to include new variables that reflect changes to the organizational structure of baseball.

There are many different ways for team to attempt to maximize their profits, but the primary driver of profit is attendance. We will assume that any individual who is willing to watch the games on TV could also attend the baseball game if they chose to. Although naming rights, television deals, and revenue sharing help increase the profitability of the league, they are typically negotiated by the teams and companies dependent on forecasted viewership and attendance.

Attendance has a direct impact on revenue with gate sales, concession and memorabilia purchases, and it can be used to measure the anticipated demand for watching the games on TV. If a stadium is selling out, it can be expected that there will be a large number of individuals watching the game on TV as well. This is where econometrics can provide an analysis of what causes consumers to care enough about games being played to pay money to attend the game.

If the best way to maximize profit is through attendance, what is the best way to maximize attendance? One way to maximize fan attendance may be through manipulating home field advantage. This hypothesis was originally tested by Knowles et al. (1992) in the context of the uncertainty of outcome hypothesis. Their test rests on the hypothesis that casual fans attend more games if each team has a chance to win but the home team is favored. When determining whether to go to a baseball game, the game will only be an attractive option if there is some amount of competition in the game (Schmidt, 2001).

An analogy is the Harlem Globetrotters. If the Globetrotters did not have a comedy act or perform tricks, fewer individuals would watch players with clearly superior abilities dominate their ever-hapless opponent, The Generals. This hypothesis is one of the key assumptions for generating fan interest. Consumers do not want to go to a game unless they are confident that there is competitive balance (Knowles et al., 1992).

As such, every team has an incentive to attempt to win games, because winning games is positively correlated with attendance and other sources of revenue (Whitney, 1988). However, winning games also increases costs in player and manager salaries, training, travel, and other expenses. Therefore, all 30 major league baseball teams will take actions in an attempt to maximize their number of wins during a season, subject to a budget constraint determined by each team's owner.

Competitive balance is one of the main points of the uncertainty of outcome hypothesis, because one team will have a greater chance to win the game if they aren't each competitive. The only way to remain competitive in baseball is to attempt to get the best players. If every team is successful in fielding teams with talented players, any team has a chance to win the game they are playing on a given day. The thought that in any game these top caliber players can win or lose is thought to be appealing to fans and can generate interest and attendance to the games (Szyamnski, 2003).

In MLB, Knowles et al. (1992) determined that there is a positive but decreasing relationship between attendance and the probability of the home team winning the game. They went on to build a model that determined that the home team should have around a 60% chance of winning the game in order to maximize attendance, all else being equal.

This indicates that a casual fan will attend a game only when the home team has a slightly more than equal chance to win the game.

This paper aims to replicate Knowles et al., and to improve on their methods. Specifically, the econometric models control for the different variation within each of the teams or between the 81 home games. Additional independent variables are included to control for inter-league play, inter-city games, over-under betting line, number of weeks in the season, the number of wins each starting pitcher has, the record of each team, and if the team is considered a rival. The final statistical results provide insight into the optimal winning probability that will maximize fan attendance in 2013, and allows us to compare to the results in Knowles et al. from 25 years earlier.

The results of this paper rely on a few key assumptions. First, and perhaps most important, is that we are considering the actions of the casual fan. Hardcore fans might attend the baseball games, despite any factors that would dissuade a casual fan. Another key assumption for this paper is that those who attend a game care more about the home team playing than the away team. Logically this makes sense, because casual fans who attend a game likely live in or near the same city as the team, and thus root for the home team. Third, it is assumed that the teams attempt to perform their best during every game, and every team's goal is to win each game they play. Finally, the betting markets are correct, and odds are reflective of the true market expectations about the game's outcome.

CHAPTER TWO: LITERATURE REVIEW

This study is directly based on Knowles, Sherony, and Haupert's 1992 paper analyzing attendance and the uncertainty of outcome hypothesis. However, many other papers have made significant contributions to the literature. This chapter reviews some of the most influential papers, highlights the results, and discusses the conclusions.

Stefan Szymanski references competitive balance and the uncertainty of outcomes in sporting events in his 2008 article "The Economic Design of Sporting Contests." He states that baseball and certain other team sports should be allowed to operate outside certain antitrust rules. These rules allow leagues maintain an even playing field and equality of resources. Inequality of resources, or players, leads to unequal competition and ultimately reduces the level of interest fans have in watching an event.

This proposition is true in MLB, and serves as the groundwork for rules regarding talent acquisition and revenue sharing. These rules attempt to prevent large market teams like New York or Los Angeles from gaining a competitive edge and reducing the competitive balance for the league. Large market teams can leverage the large populations of their host cities to generate more revenue, and then spend that extra revenue to acquire the best talent and build the best teams so that they win perennially. If these rules are enforced, and achieve the expected effect, the competitive edge that large market teams have will decrease and the games played will be competitive (Szymanski, 2008). Games that are competitive include two teams that have a similar chance to win each game, and it is hard to predict the winner with certainty. Uncertainty generates interest by fans because they don't know the outcome of the games being played, and want to watch the team they root for win those games.

Perhaps one of the most cited papers stating there is a relationship between the uncertainty of outcome hypothesis and attendance in Major League Baseball is written by Knowles, Sherony, and Haupert (1992). Their paper uses betting odds in baseball as a proxy for the pre-game subjective probability of the home team winning. Knowles et al. hypothesize that there is a positive and diminishing return to the subjective win probability of the home team. Their empirical analysis uses data from the National League during the 1988 season. The model includes a number of factors to control for other influences on baseball attendance.

The results of their analysis predict that attendance in 1988 was maximized when the home team is slightly favored to win the game; their optimal winning probability is close to 0.6 (Knowles et al., 1992). Although the results of the study are statistically significant and are consistent with economic theory, their paper does not include any baseball related variables other than the combined games back of each team.

Rather their statistical model focuses on different factors that influence the choices of consumers, such as income and if the game was played on a weekend or evening, but it does not include any extra variables such as the records of the starting pitchers, when in the season the game is played, or account for whether the game was expected to be high scoring. All of those variables potentially influence fan interest, and could diminish the effect the uncertainty of outcome on attendance.

Another paper that provides insight into the relationship between attendance and the probability of winning is written by Schmidt and Berri (2001). Their paper states that in order for the MLB to maintain some form of competitive balance the level of talent disparity across the teams would have to be reduced. Once competitive balance is achieved, there will be a positive relationship between it, and attendance (Schmidt and Berri, 2001).

Schmidt and Berri’s paper improves upon the previous literature by looking at the affect marginal wins have on attendance (Schmidt, 2001). Their first hypothesis states that teams attempt to maximize the number of wins by acquiring talented players. During 1990-1999, there was a large degree of disparity between teams, which signified the league was not competitive; most evident from the Yankees and Braves success during this era (Schmidt and Berri, 2001). Using a Gini coefficient, Schmidt and Berri show that the competitive balance during this time was higher in 1990 than during the rest of the 20th century. Finally, Schmidt and Berri compare the Gini coefficients to attendance as a test for whether competitive balance impacts attendance and fan interest. They find a positive relationship between the Gini coefficient, which measures competitive balance, and attendance for the one year model. This means that as the league became less competitive, attendance increased.

Their other models compare the level of competitive balance over three and five year periods and find a negative relationship with attendance (Schmidt and Berri, 2001). Their results confirm that there is a positive relationship between competitive balance and attendance over a set period of time. It is possible that in the short term, fans attend games to see very dominant teams playing and winning against poor teams. Schmidt and

Berry theorize that this was the case during the 1990s when the New York Yankees and Atlanta Braves exhibited a strong degree of dominance over their opponents.

There is also a wide range of economic literature that contradicts Knowles et al.. One of the most convincing papers was written by Buraimo and Simmons in 2008, who use the same framework as Knowles et al. to analyze the effects of attendance in the English Premier League. The EPL, in contrast to MLB, does not impose any rules that limit the monopolization of talent balance between the teams (Buraimo and Simmons, 2008). This laissez faire approach reduces the amount of competitive balance in games, and thus would be expected to affect the relationship between the uncertainty of outcome hypothesis and attendance. Buraimo and Simmons use data from the 2000-2001 English Premier League (EPL) season and find that the relationship is actually the opposite of Knowles et al. (Buraimo and Simmons, 2008). The study finds that fans of the EPL actually prefer to attend the games where there is a considerable favorite i.e. when there is little uncertainty in the outcome (Buraimo and Simmons, 2008). The reasons offered by Buraimo and Simmons is that EPL fans would rather see their team dominate a worse team then watch an evenly matched game in which the home team stands a significant chance to lose (Buraimo and Simmons, 2008).

Buraimo and Simmons (2009) explain why they find the opposite relationship from a large bulk of economic literature on the uncertainty of outcome hypothesis. The paper primarily deals with four years of data from Spanish football and uses the same methodology as their earlier paper, but separates the audience between those who watch games on TV and those who attend the match live. Casual fans are likely to watch the game on TV; hardcore fans care enough about the team to buy tickets to go see the game live. Buraimo and Simmons find the same relationship as Knowles et al. between the uncertainty of outcome and attendance exists with the fans who watch the game on TV, but the reverse for the fans who attend the games (Buraimo and Simmons, 2009).

An interesting interpretation about the increased demand for less uncertainty is that fans at games are looking for their team to put on a show and demonstrate their superior abilities. Being able to watch the players demonstrate their superior ability and competitive advantage over the opponent will typically create an exciting and high scoring game. A conclusion taken from Buraimo and Simmons is that that fans will decide to attend the game based on how exciting they expect the game to be.

The Buraimo and Simmons (2008, 2009) papers are useful to consider and bring a different point of view to the literature. Their papers show that the relationships between attendance and competitive balance and the uncertainty of outcome rely on the assumptions and definitions of each model and analysis. The difference between the two papers is that the second one accounts for the fact that different types of consumers react differently based on the chance the home team has of winning the game.

Berkowitz, Depken, and Wilson (2011) also found this effect in their paper studying NASCAR. They show that NASCAR attendance is segmented between those who attend races and those who watch on TV. Fans who attend races are typically not affected by factors that influence TV viewership. More people watch the races later in the season or if the race is expected to be competitive. People who watch the races on TV have considerably more substitutes during the race, and require a very competitive race with a high degree of uncertainty in the outcome to stay interested in the race.

It is important to note that many different factors other than betting odds go into measuring the uncertainty of outcome in baseball games Forrest and Simmons (2002) compare attendance to both uncertainty of outcome and the competitiveness of the teams playing the game. Their main contribution is the inclusion of the assumption that the odds are biased and do not accurately predict which team will win the game. The odds are influenced by many other factors than just how competitive the teams are (Forrest and Simmons, 2002).

Forrest and Simmons focused on the English Football League, because the odds are set days in advance and do not change based on factors that might occur between when they are set and when the game is played. Finally they account for the potential bias created by inefficiencies in the betting market, along with the probability of a draw occurring. The results show that most consumers are drawn to games where there is no strong competitive advantage that a team has over the other (Forrest and Simmons, 2002).

All of the papers listed above show the diversity in relating the uncertainty of outcome, or competitive balance, to attendance. These articles can help display the overarching theory supporing the importance of this topic; however, there are additional when determining the model.

Michael Butler hypothesized that the implementation in interleague play increased revenue by approximately 7% (Butler, 2002). Butler's paper models the impact interleague games have on attendance, and although we can make an argument that the teams have some degree of competitive balance, the interleague games might negatively impact the relationship between the uncertainty of outcome hypothesis and attendance.

The reason behind this is that the home team is playing a new opponent that they usually don't play, which would generate extra excitement and fans would not be as concerned with who will win the game. The interleague play allows certain match ups between teams in or near the same metropolitan area.

Other research has shown that in order to increase the chance of winning a game teams must be able to afford to pay free agents money during free agency. This money can help sign superstars who help provide teams a competitive advantage due to their superior skills (Rivers, 2002). Rivers also points out that just having a superstar on your roster can increase attendance (Rivers, 2002). It would be logical to assume that this increase in attendance can work both ways. If a visiting team has a superstar on their team, it is more likely that that game will attract fans to see that superstar play.

A useful corollary would be fans of competing NBA teams might go to home games to see Michael Jordan play. This would also have a negative effect on the relationship between the uncertainty of outcome hypothesis and attendance. Unfortunately superstar players aren't available to every team, and need to be acquired through trades or spending large amounts of money in free agency. The need for teams to recruit superstars to gain some form of advantage over other teams automatically gives large market teams an advantage because they have more money from revenue and can afford spending larger amounts of money (Walker, 1986). The size of each team's market gives an advantage to large markets and decreases competitive balance. Since the uncertainty of outcome hypothesis relies on competitive balance it might not have as large an effect on attendance.

CHAPTER THREE: METHODOLOGY

This chapter describes the data used in the paper, and addresses the specific statistics and modeling performed. The research techniques in this paper consist of a quantitative analysis of Major League Baseball during the 2013 season. The benefit of using MLB compared to other leagues is that the full season has 162 games for each team. This helps create a wide range of variability in the data and has frequent match ups of different teams that will repeat over the course of the season.

The data describe all 81 home games for each of the 30 teams, for a total of 2430 observations. The away games from each team are not considered because the attendance at each game is composed of fans who live in the city the game is being played in; fans in attendance are primarily fans of the home team. Cities with multiple teams or other MLB teams nearby can cause the number of fans of the home team to decrease. The sample can be representative of the targeted population of any season from the MLB because fans will usually not change their favorite team between seasons, and the schedule keeps the same number of games played and with the exception of interleague play, the opponents are fairly consistent throughout the season. It would take a structural change to the league to make the data set not be representative of that year. Examples of structural changes to the league would include teams moving or changing cities, expansion or contraction of teams, changes to the schedule, shuffling of the divisions, and possible strikes shortening seasons. A benefit of using data from 2013 is that interleague play has increased in each season and now features a record high number of games played between the two leagues.

The data were collected using different independent websites, and verified with the data on ESPN.com and MLB.com. This was done in an effort to ensure that the independent data sources are accurate. Once all of the relevant baseball statistics were collected, the subjective probabilities that the home team will win were calculated using historical odds set up by an online sports book.

These odds, called money lines, are determined before the games were played. Money lines are used in games where the final margin of runs does not vary by much or games are typically low scoring. This is especially true in baseball, soccer, and hockey; while other sports where margins of victory can be substantially higher it makes more sense to use a point spread method. Money lines are expressed for the two teams as a positive number for the underdog, and a negative number for the favorite. The scale of most money lines is how much money is needed to gain $100 for the favorite, or how much money can be gained by betting $100 on the underdog. Using Equations 1 and 2, the money lines can be easily converted to the subjective probabilities each team has to win the game.

$$\text{Subjective probability of favorite} = \frac{(\mathrm{L_f}+\mathrm{L_u})/2}{((\mathrm{L_f}+\mathrm{L_u})/2)-100} \tag{1}$$

$$\text{Subjective probability of underdog} = \frac{100}{((\mathrm{L_f}+\mathrm{L_u})/2)+100} \tag{2}$$

Where: $L_f \equiv$ the money line for the favorite. Expressed as a negative number

$L_u \equiv$ the money line for the underdog. Expressed as a positive number

Equations 1 and 2 calculate the subjective probability the home team has to either win or lose the game depending on if the home team is a favorite or underdog. These equations take the average between the absolute values of each money line. This normalizes the difference between the two different values, and causes the odds for each team to sum to one. The average between the home and away team's money lines must be taken because of the spread each casino or sports book uses. Sport books build a margin into their odds to help hedge against losing money depending on the number of bets. The money the sport book is expected to pay out depending on the winner is covered by the number of bets taken on the loser. If the number of bets on the loser does not cover the amount paid to the people who bet on the winner, the sport book will lose money. Because of this, the odds are set up to influence betting behavior and hedge against the sport book losing money, and the summed calculated subjective winning probability will not equal one. Taking the average of the absolute value of the money lines will control for the sport book's hedge against losing money.

There are a few assumptions about the money lines and their implementation moving forward. The first and most important is that consumers either know the subjective probabilities, or already have a good idea about which team is better and favored to win. The second is that the betting market is fair, and accurately lists the money lines with what is expected to occur during the game.

Once all the data were collected and checked for accuracy and subjective probabilities calculated, additional data needed to be generated. The win and loss records from the starting pitchers, week and month of the season, day of the week, and dummy variable for whether the game was a weekend, and dummy variables identifying the league, and division, interleague games, intracity games, and rival games were created. Once all of the extra variables had been created, they can be used to modify Knowles' et al. original model and equation. The modifications are added in an attempt to explain any additional variation that was not captured by Knowles et al. in their paper.

The Knowles' et al. model is replicated using current data in order to insure that the relationship still holds after 25 years. During that time, there have been multiple changes to the league including the addition of interleague play, expansion teams, and relocated teams. The replicated model will determine if any of these changes had an effect on the relationship between attendance the uncertainty of outcome hypothesis.

The replicated model is then modified to account for a number of factors that Knowles et al. do not control for in their paper. This modified model includes variables that control for interleague play, the over/ under betting line, rival opponents, each team's record, the wins of each pitcher, and season week in which the game is played. The modified model also uses panel data models to control for variation among the teams as well as for variation occurring across the 81 home games. This modified model is then compared to the replicated model to determine if any of these variables changes the relationship between the subjective probability of the home team winning and attendance. If any of the new variables have a significant effect on attendance, they might negate the affect the uncertainty of outcome hypothesis has on attendance.

The modified models are estimated using various methods including pooled OLS, and panel data estimates which account for fixed and random effects. The pooled OLS, fixed effects, and random effects models attempt to compare the results not only over time but also across the different teams. Using panel methods it is possible to obtain different coefficients and levels of significance for the independent variables, and control for different variation within teams and across the games played. This variation is not controlled for in the OLS or OLS with clustered standard errors. Based on the results of the modified models we can determine if the results of Knowles et al. are still applicable and if attendance is driven by the uncertainty of the home team winning each game. The scope of the results will show if any of the extra variables and econometric techniques provide a different outcome, or further confirm the findings from Knowles et al. (1992).

CHAPTER FOUR: DATA ANALYSIS

The previous section addressed the motivation for this study; this chapter will explain how we will test our hypothesis. This includes a description of where the data were obtained, the final analysis and econometric methods. This empirical analysis attempts to determine if a relationship exists between the probability of a baseball team winning and attendance by replication of Knowles et al., then will improve upon their original methodology with new econometric techniques and including extra independent variables. The main advantage of replicating the Knowles et al. study is to update the data to ensure that the relationship holds up after more than 20 years since the paper was published. This is important due to changes made to the league since the 1988 season are structural. All data used in this empirical analysis were obtained from different sources including baseball-reference.com, covers.com, and espn.com. The data set was then parsed for missing, truncated, or incorrect values; then checked against each team's website and mlb.com to ensure validity and accuracy. The variables used in the data set are listed in Table 1, while descriptive statistics can be found in Table 2.

TABLE 1: Coefficients and descriptions	
Coefficients	Description
attendance	Number of tickets sold at each home game
winprob	Subjective probability of the home team winning
winprob2	Squared subjective probability of the home team winning
gb	Number of games back the home team is from first place in their division
weekend	Dummy variable if the game was played on the weekend. 1 if true
pop	Population of the city the stadium is in
urate	Unemployment rate of the city the stadium is in
income	Median per capita income of the city the stadium is in
dist	Distance, by air, between the cities of the teams playing
recordl1	The number of wins the home team has entering the game
pitcherwin	Number of wins the current pitcher has entering the game
double	Dummy variable for if the game is the second game in a doubleheader. 1 if true
week	Number of week in the season
streak	Number of consecutive wins and losses the home team has entering this game
intracity	Dummy variable for if the teams playing the game have stadiums in the same city. 1 if true
interleague	Dummy variable for if the teams are playing an interleague game. 1 if true
rival	Dummy variable for if the team is playing a rival team. 1 if true
overunder	The expected number of total runs scored between both teams

TABLE 2: Descriptive statistics of coefficients					
Coefficients	Count	Mean	Standard deviation	Minimum	Maximum
attendance	2430	30508.710	9802.800	9143.000	53393.000
winprob	2430	0.542	0.083	0.283	0.769
winprob2	2430	0.300	0.090	0.080	0.591
gb	2430	14.802	11.919	0.000	58.000
weekend	2430	0.319	0.466	0.000	1.000
pop	2430	1613513.000	2069576.000	296550.000	8336697.000
urate	2430	5.347	0.849	4.000	7.000
income	2430	47470.970	12349.660	26217.000	75604.000
dist	2430	1091.365	1077.481	0.000	11922.700
recordl1	2430	21.292	13.341	0.000	55.000
pitcherwin	2430	4.185	4.852	0.000	25.000
double	2430	0.011	0.103	0.000	1.000
week	2430	13.751	7.541	1.000	27.000
streak	2430	0.203	2.538	-15.000	14.000
intracity	2430	0.003	0.057	0.000	1.000
interleague	2430	0.123	0.329	0.000	1.000
rival	2430	0.272	0.445	0.000	1.000
overunder	2430	8.018	0.878	6.000	11.500

Once the data were verified, they were analyzed for autocorrelation, heteroskedasticity, multicollinearity, and non-stationary in the variance or mean. Without any of these typical problems present, and the data's accuracy already checked, any doubts about the quality of the data can be easily dismissed. The absence of these issues also lends credibility to the significance of the analysis because the analysis is not biased, and the results accurately reflect the true population parameters. This will provide some level of internal and external validity. In the absence of any underlying issues, the data can be used in a variety of different models. A complete list of these models is found in Table 3.

TABLE 3: List of models	
Model	Description
1	Pooled OLS
2	Pooled OLS with clustered standard errors
3	Panel Fixed Effects
4	Panel Random Effects
5	Panel Between Effects

The pooled OLS regression model is similar to a linear regression model, and is used by many of the papers covered in the literature review. The analysis of this model is adequate, but the model can leave out several omitted variables. A pooled OLS regression requires few assumptions, and everything that is not controlled for is aggregated into the error term. The main benefit of this model is the ability to combine all the teams and games played for a total of 2,430 observations. The main cost is potential bias if there is unobserved heterogeneity across teams in MLB, i.e. teams behave differently based on the different levels of talent they have, or certain stadiums favor hitters or pitchers based on their characteristics.

The second model uses the same equations as the pooled OLS model, but controls for possible correlation within teams when calculating standard errors. This is important because the model can control for differences in weather in each city, or the peculiarities of each ball park that could cause heteroskedasticity as the season progresses. The other three models are variations of panel estimators. The panel models treat each of the 30 MLB teams as a unique unit, and treats the

81 home games each team plays as a time period. The fixed, random, and between estimators use different types of variation in the data. Fixed effects models use the variation within the team and takes advantage of any variation within each of the 30 teams. Between effects estimators use the variation between teams by taking the mean of each variable through the 81 games across each team. The random effects model works similarly to the fixed effects model, but instead of effectively adding a dummy variable for each team, the random effects model allows the possible heterogeneity among the groups and time to be distributed normally across the entire population. This method is effective when the data constitutes a sample of the overall population.

Once the proper model is selected that accounts for all of these alternative types of variation, and the assumptions of the model does not limit the analysis, we can estimate the parameters of the different models using three unique equations. These equations are listed in Table 4. Each equation is used for a different purpose. The first equation tests whether there is a relationship between the probability of the home team winning and attendance at the game. The second equation is a replication of Knowles et al.'s original model. The third equation includes different independent variables not used in any previous study on the uncertainty of outcome hypothesis and attendance. The inclusion of these variables is meaningful in explaining what drives attendance, and if attendance is driven by the probability of the home team winning.

TABLE 4: Independent variables and equations	
Relationship	Formula
Naïve	attendance = β1*winprob + β2*winprob2 + u
Knowles	attendance = β1*winprob + β2*winprob2 + β3*gb + β4*weekend + β5*pop + β6*urate + β7*income + β8*dist + u
Modified	attendance = β1*winprob + β2*winprob2 + β3*gb + β4*weekend + β5*pop + β6*urate + β7*income + β8*dist + β9*recordl1 + β10*pitcherwin + β11*double + β12*day + β13*week + β14*streak + β15*intracity + β16*interleague + β17*rival + β18*overunder + u

The parameters in each of the equations in Table 4 are estimated using each estimator in Table 3, which provides a total of 15 different results to evaluate. Using intuition, economic theory, and the econometric results these different models and equations can be judged to determine if any of the models accurately predicts the true relationship between the probability of winning and attendance.

Once the estimated coefficients are calculated the final step is to calculate the optimal winning percentage that will maximize attendance. This is accomplished using Equation 3:

$$Prob_{Max} = \frac{\beta_{winprob}}{-2 * \beta_{winprob^2}} \tag{3}$$

Where: $\beta_{winprob}$ $\equiv$ the probability of the home team winning

$\beta_{winprob^2}$ $\equiv$ the probability of the home team winning squared

The resulting percentage is home team’s winning percentage that will maximize attendance, controlling for all the other independent variables, variation within and between the teams and games, and other things held constant.

CHAPTER FIVE: FINDINGS

The chief concern of this paper is if the Knowles et al. study can be replicated to see if the relationship between attendance and the uncertainty of outcome exists then improve Knowles' et al. research by including additional variables and different models. The data used in Knowles et al. was from the 1988 season; significant structural changes occurred in MLB since then. Four MLB teams were added, two teams moved to different cities, two divisions were added, and interleague play was added to include regular season games between the American and National Leagues. These structural changes may have diminished the relationship between attendance and the uncertainty of outcome hypothesis. This would cause the optimal winning percentage of the home team be different from the finding by Knowles et al.. The results will be presented in three sections. The first will discuss the findings. The second part will discuss differences between these results and those presented in Knowles et al.. The final part will explain the implications of the results.

Before an in depth discussion of the statistical findings, it is helpful to investigate a number of graphs that can help explain and support the different results given by the models. Recall that the data set used in this paper is comprised of home games for each of the 30 MLB teams for the 2013 season. The probability of the home team winning each game for all 30 teams can be seen in Figure 1. This set of graphs reveals there is a moderate to large degree of variation in the expectation of the home team winning. No team consistently is expected to win or lose during the 81 home games. These graphs help show that overall no one team seems to be favored to win more than the others, and might imply that competitive balance among the league is fairly high.

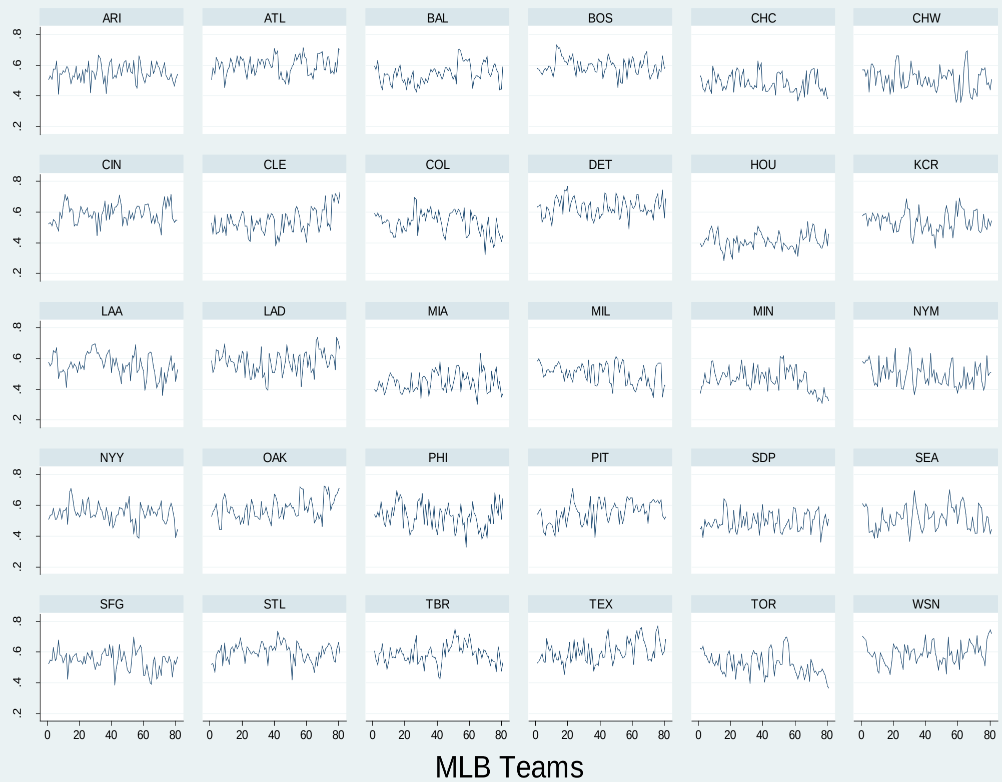

FIGURE 1: Winning probability of all 30 MLB teams

The assumption of competitive balance is important because it means at any given point in the season any team has a chance to beat any other team. This drives the uncertainty of outcome hypothesis. If either the home or away team can win it helps create an interesting game for consumers. Knowles et al. pointed out that although the uncertainty of outcome hypothesis draws individuals to attend home games, they also prefer attending games in which the home team has a chance to win. The next step would be to use a “naïve” approach and find whether a relationship between the probability of winning and attendance exists. Figure 2 and 3 can help with this.

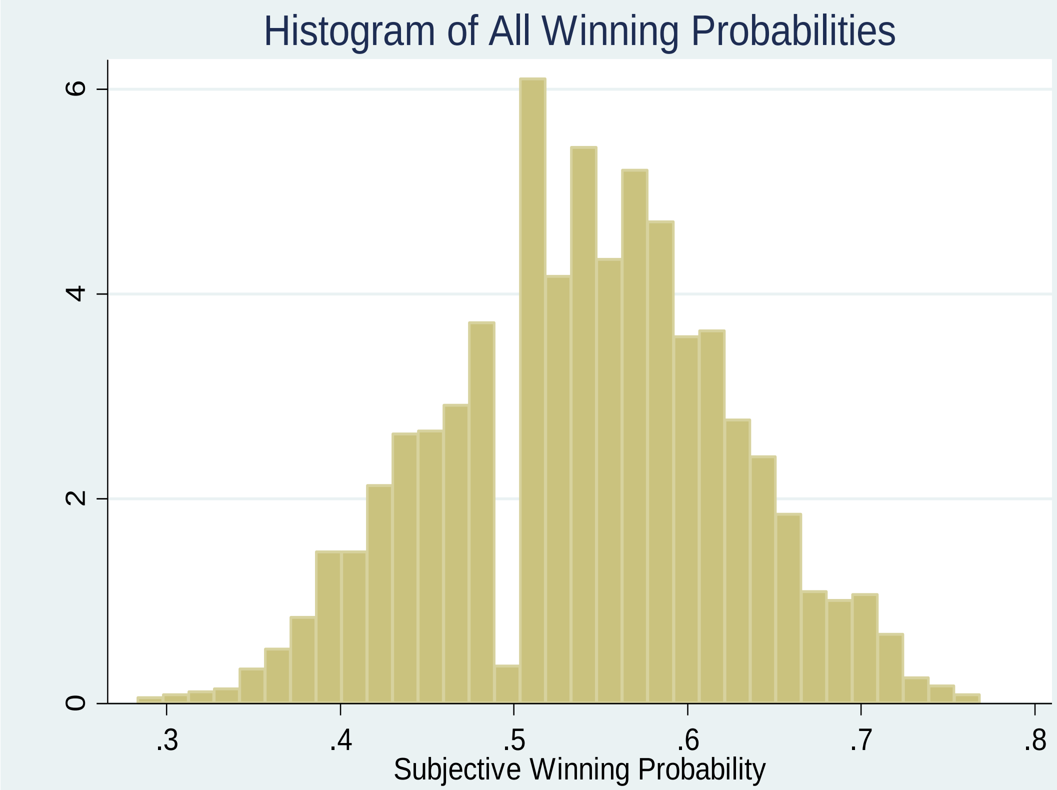

Figure 2

Figure 2 is a histogram of the different subjective winning probabilities that the home teams faced. The probabilities are centered at 0.54 and show that home teams experience a meaningful home field advantage in the betting markets. The absence of games with a 0.5 probability demonstrates that sports books and bookies will typically not assign “pick ‘em” status to games. Typically it is assumed that the odds makers know something special about a team that causes them to set the odds to be different from a 50-50 split. Typically the team’s home field advantage will give a slight edge to the home team, and cause the betting to swing in favor of them. Because of this, we can expect the bulk of the winning probabilities to be just over the 0.5 mark.

Knowles et al. estimated that the optimal winning percentage is 0.6, which is close to where the bulk of the data on the histogram lie. This graph adds credibility to their claim. The graph does not control for the different scenarios where the home team had a subjective probability different from 0.6, but still drew a large attendance. Different factors such as the popularity of the opposing team, a superstar player visiting, special promotions like fireworks and giveaways, and many more examples can influence attendance independent of the expected outcome of the game, and why modifications of the Knowles et al. model is needed.

We can also use Figure 3 to explain this. Figure 3 graphs attendance of each team throughout the season. This graph shows that there is considerable variation across each team's home games, except the San Francisco Giants (SFG), who are almost always at capacity. If the bulk of the winning probabilities fell close to the optimal winning percentage, we would expect to see more of the graphs resembling the SFG subplot.

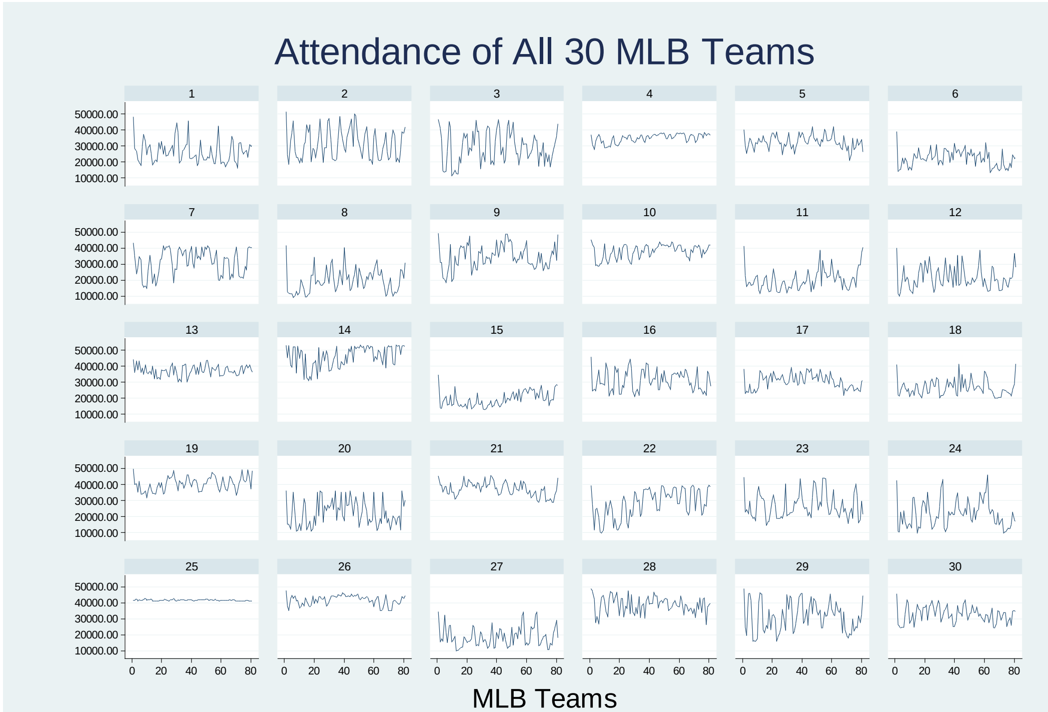

Figure 3

The final and perhaps most important figure that can add to the argument that the probability of winning can influence attendance is displayed in Figure 4. This figure shows the relationship between winning percentage and attendance, then graphs a quadratic function through those points. This quadratic equation is the same as the naïve model, which can be found in Table 3, in Chapter 5. This figure shows a positive, but decreasing relationship between the probability of winning and attendance.

This relationship reinforces the previous findings of Knowles et al. and shows one of the many factors influencing attendance is the probability of the home team winning. There are several takeaways that can tell a different story. The quadratic reveal a relatively inelastic relationship suggesting a change in the winning probability will not cause a large change in attendance at games. Using the naïve model, a quadratic is depicted on this graph, but there appears to be a weak relationship as the bulk of the data points do not fall close to the quadratic line. This also shows that there is a possibility that the relationship between the probability of winning and attendance are not as strong as suggested by Knowles et al., or a different model should be specified that controls for the variance of the data. The earlier explanation why no points are at 0.5 apply to this figure as well.

Overall variation: Attendance vs Winprob

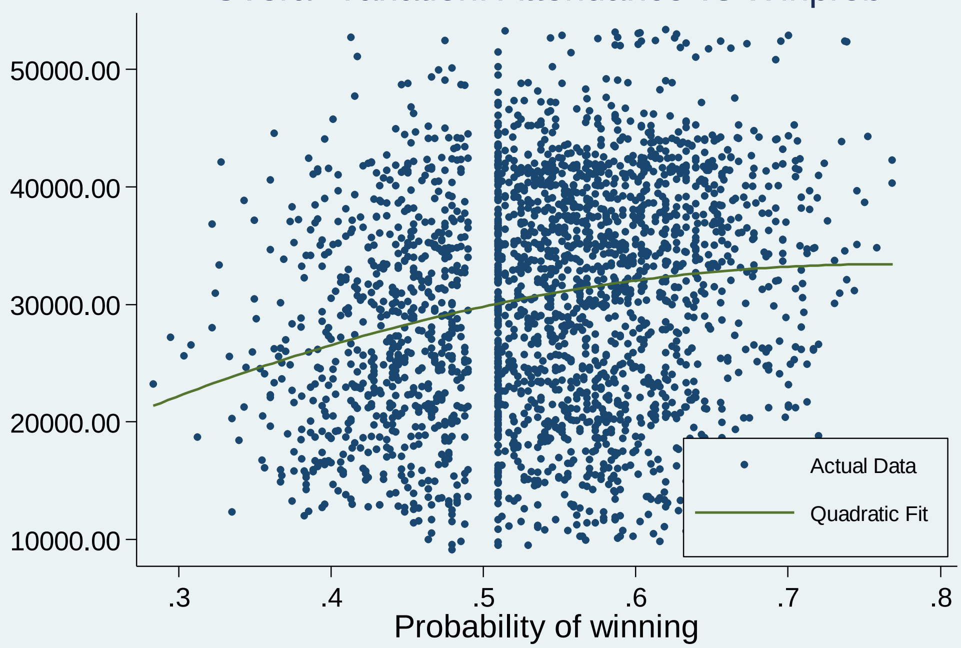

Figure 4

All of the figures presented tell a slightly different story, but are important in their own respect. They each point to different facets of the problem, and can add credibility to some of the hypotheses in the literature. The figures also point to different stories within the literature review. Because of this, we can’t rely solely on them and must turn to the conditional models. The resulting discussion talks in depth about the quantitative results and their interpretation.

The econometric results are separated into three tables in the next few paragraphs. Each table reports the estimated coefficients, and standard errors below them in parentheses. The estimated coefficients are also marked with stars to represent testing statistically significant at different levels of significance. Three stars (***) denote coefficients that are significant at the 0.01 level, two stars (**) are significant at the 0.05 level, and one star (*) is significant at the 0.1 level. Any of the other coefficients did not test statistically significant and are not different from 0.

There were five econometric models estimated for three different equations to determine first if a relationship exists between the probability of winning and attendance. The first model is a naïve model, and its results are reported in Table 5. The main point of this equation is to show there is a strong relationship between the wining probability and attendance using pooled OLS. The other models run did not show a strong relationship like the OLS model, and are not useful in explaining any relationship between the two variables.

TABLE 5: Naïve econometric model estimated coefficients and standard errors

Variables	OLS	Cluster	FE	RE	BE
winprob	80173.270***	80173.270*	8544.201	9790.212	66170.173
	(23646.233)	(40523.119)	(16997.308)	(19257.030)	(469062.728)
winprob2	-52646.973**	-52646.973	-9014.568	-9804.437	2437.094
	(21903.958)	(36178.032)	(15555.638)	(17618.651)	(444491.722)
Intercept	2895.847	2895.847	28588.056***	28150.401***	-6060.425
	(6298.228)	(11383.252)	(4597.489)	(5541.681)	(121003.727)
R^2	0.043	0.043	<0.001		0.208
F stat	54.035	4.572	0.375		3.553

Legend: $p < 0.1$ = * $p < 0.05$ = ** $p < 0.01$ = ***
Coefficients / standard errors

The second equation was simply a recreation of the model estimated in Knowles et al. (1992). This equation was estimated to see if the relationship estimated in their paper holds during a different time period, after substantial structural changes made to the scheduling of baseball games, expansion teams, and the relocation of other teams. The results for this model are reported in Table 6.

The OLS results for this equation replicate Knowles et al. with the proper sign and significance for all of the variables except for the winning probabilities. Even using the different panel models, a large number of coefficients test significant, and take on the same sign and magnitude as the traditional OLS model. Knowles' et al. findings are robust because they persist through multiple time periods, and structural changes in MLB. The problem with this model is that the coefficients for probability of winning and its quadratic are both insignificant. The model did not show a positive, significant relationship between the betting markets odds of the team winning, and attendance.

TABLE 6: Knowles econometric model estimated coefficients and standard errors and standard errors

Variables	Knowles et al.	OLS	Cluster	FE	RE	BE
gb	-293.238***	-159.594***	-159.594***	-45.67***	-47.477***	-175.521
	(29.427)	(15.561)	(38.996)	(11.736)	(15.752)	(578.800)
weekend	9026.45***	5119.586***	5119.586***	5145.99***	5145.495***	137278.679
	(794.239)	(383.518)	(663.489)	(251.854)	(642.945)	(191510.018)
day	3415.12***	3080.59***	3080.59***	3413.972***	3410.34***	-9542.305
	-828.834	(379.258)	(609.250)	(252.455)	(506.220)	(17952.639)
pop	0.00162351***	0.001***	0.001	(omitted)	0.001	0.001
	(0.001)	0.000	(0.001)		(0.001)	(0.001)
urate	28020.300	874.17***	874.170	(omitted)	1214.018	335.207
	-42252.200	(250.038)	(1416.918)		(1703.806)	(2172.488)
income	0.208	0.108***	0.108	(omitted)	0.115	0.189
	(0.158)	(0.016)	(0.094)		(0.111)	(0.139)
dist	-1.85896***	-0.522***	-0.522*	-0.142	-0.147*	-5.204
	(0.389)	(0.167)	(0.266)	(0.113)	(0.084)	(5.318)
winprob	214976**	-11453.446	-11453.446	-4865.297	-4648.124	-429581.798
	(107782.000)	(22617.106)	(32319.030)	(15543.175)	(17787.169)	(632914.607)
winprob2	-179211**	34292.281	34292.281	5399.016	5551.793	457182.930
	(101170.000)	(20935.440)	(32425.964)	(14318.197)	(15946.016)	(576894.333)
Intercept	-47103.4*	15724.883***	15724.883	29566.656***	16738.644	81004.704
	(28840.500)	-6022.936	(9978.323)	(4201.811)	(10362.953)	(181339.290)
R^2	0.376	0.207	0.207	0.190		0.387
F stat	56.980	70.102	14.349	93.360		1.406

Legend: $p < 0.1$ = * $p < 0.05$ = ** $p < 0.01$ = ***
Coefficients / standard errors

Finally, the modified model's coefficients are listed in Table 7. This model adds variables that account for the week of the season, intracity games, interleague games, if the game was a doubleheader, the current win or loss streak the team is on, if the opponent is considered a rival, and the betting market's over/ under on the number of expected runs scored. Each of these variables attempts to add explanatory value to the original framework proposed by Knowles et al.. All of the variables except for doubleheader games, and the win or loss streak tested positive at some level of significance. This shows that the modified model accounts for different factors that were not explained or accounted for in the original framework. The statistical significance of these extra variables increases the explanatory value of what influences and impacts attendance. The probability of winning and its square however remain statistically insignificant. Tables 6, and 7 both show the same problem; the winprob variable is not significant.

TABLE 7: Modified econometric model estimated coefficients and standard errors					
Variables	OLS	Cluster	FE	RE	BE
recordl1	-6.111	-6.111	-9.729	-5.837	-2462.446
	(60.367)	(167.849)	(48.375)	(100.411)	(2115.304)
pitcherwin	32.751	32.751	-23.554	-23.066	-445.839
	(36.798)	(57.554)	(25.043)	(24.701)	(1243.290)
gb	-295.609***	-295.609***	-169.021***	-171.768***	-636.996
	(24.304)	(70.464)	(18.258)	(28.646)	(756.381)
doubleheader	414.568	414.568	101.579	107.952	172922.402
	(1674.140)	(1506.349)	(1109.234)	(1130.407)	(158622.739)
weekend	5304.923***	5304.923***	5320.964***	5317.938***	186149.731
	(374.397)	(654.983)	(247.560)	(637.439)	(195156.184)
day	3232.973***	3232.973***	3543.27***	3539.076***	15011.036
	(368.357)	(591.578)	(245.985)	(505.433)	(20259.114)
week	355.713***	355.713	249.651****	245.842	-1611.597
	(118.971)	(321.742)	(90.498)	(177.667)	(3803.259)
streak	-88.224	-88.224	-71.956	-71.954	1305.790
	(72.188)	(76.429)	(47.703)	(63.063)	(8654.630)
intracity	6930.832**	6930.832**	5177.018**	5194.302***	1070930.978**
	(3089.570)	(2671.299)	(2042.139)	(893.992)	(454128.264)
interleague	922.764*	922.764	1088.533***	1084.492**	(omitted)
	(536.386)	(594.561)	(353.664)	(540.484)	
rival	1610.676***	1610.676***	1365.166***	1365.807***	100895.476**
	(413.002)	(565.982)	(274.749)	(372.184)	(36520.209)
pop	0.001***	0.001	(omitted)	0.001	-0.004*
	0.000	(0.001)		(0.001)	(0.002)
dist	-0.190	-0.190	0.022	0.018	-12.882
	(0.167)	(0.282)	(0.114)	(0.087)	(7.481)
income	0.109***	0.109	(omitted)	0.108	0.118
	(0.015)	(0.091)		(0.101)	(0.150)
urate	971.963***	971.963	(omitted)	1092.153	-251.207
	(244.633)	(1360.878)		(1530.840)	(3145.456)
overunder	603.638***	603.638	-205.615	-189.690	-3669.729
	(201.826)	(867.360)	(192.606)	(163.713)	(3049.578)
winprob	-18328.619	-18328.619	3676.636	3760.551	-132569.811
	(22076.568)	(30086.252)	(15180.653)	(16567.798)	(617277.321)
winprob2	42071.868**	42071.868	4618.962	4870.110	220688.113
	(20352.835)	(30144.250)	(13966.114)	(15200.162)	(559003.824)
Intercept	7692.616	7692.616	24734.113***	12608.313	69230.806
	(6087.820)	(10770.142)	(4294.486)	(9494.315)	(179941.753)
R^2	0.218	0.218	0.215		0.690
F stat	39.650	14.713	46.785		1.807
		Legend:	p < 0.1 = *	p < 0.05 = **	p < 0.01 = ***

Coefficients / standard errors

The different models presented in the tables above all have been judged by some criteria to determine which of the five models, the naïve, Knowles, and modified models, are best. The applicable p-values are in Table 8. The F-test tests the fixed effects panel data and we reject the null hypothesis that the pooled OLS provides a better model than fixed effects (FE) for every model. Breusch-Pagen test of random effects is used to determine if the random effects model is better than Pooled OLS. All of the models rejected the null hypothesis for this test as well. The Hausman test is used to compare the fixed and random effects models. Only the naïve model rejected the null hypothesis that the two techniques were the same. The random effects model is superior to every other technique used in this paper.

TABLE 8: P-values of models

	Naïve	Knowles	Modified
F-test	<0.001	<0.001	<0.001
Breusch-Pagen	<0.001	<0.001	<0.001
Hausman test	0.004	0.27	0.147

A number of different factors help to explain the lack of significance for winprob and winprob2. First, the odds used to calculate the optimal winning percentage in Knowles et al. were the “Eastern line odds” these odds do not appear to be calculated any differently from a basic money line, except their scale is based on $5 instead of $100 (Knowles et al. 1992). No reference to these special odds could be found. It is possible that these odds are calculated differently and could produce a different set of outcomes than traditional money lines. These special line odds might provide different results that better match the results in the paper by Knowles et al.. Another plausible explanation is that consumers are less interested in the game. Baseball games have a social element beyond who is playing and who wins. Consumers might attend a game for simple entertainment and not care about the game at all; they just want to attend an activity and have fun regardless of what happens on the field. Joint tests of significance were also performed on winprob and winprob2 to determine if together they are statistically significant. The p-values for that test are in Table 9.

TABLE 9: Joint test of significance of winprob and winprob2		
Model		P-value
Naïve	OLS	<0.001
	Cluster	0.019
	FE	0.688
	RE	0.823
	BE	0.043
Kowles	OLS	<0.001
	Cluster	0.005
	FE	0.814
	RE	0.831
	BE	0.342
Modified	OLS	<0.001
	Cluster	0.058
	FE	0.659
	RE	0.733
	BE	0.195

The estimated coefficients for the probability of the home team winning are not significant, but when using Equation 3 to determine the optimal winning percentage it is significant at the 0.01 level and very close to the average of the calculated winning percentages of 0.54. The statistics are listed below in Table 10. The calculated probability is equal to 0.5203. The 95% confidence interval includes Knowles, Sherony and Haupert's estimate of 0.6. This means that there is a relationship between attendance and the uncertainty of outcome hypothesis. However, the 95% confidence interval spans from 0.31 to 0.74, which is a substantially large interval. Because this variable has so much variation, it may not be one of the factors in explaining attendance at MLB games. The modified model finds there are other variables that explain attendance better than the replicated Knowles et al. study. These variables show that MLB team owners should not just concern themselves with the competitive balance of their team, in order to maximize their attendance.

TABLE 10: Optimal subjective winning probability					
Model		Coefficient	Standard Error	P-value	95% confidence interval
Naïve	OLS	0.761***	0.096	<0.001	(0.573, 0.950)
	Cluster	0.761***	0.168	<0.001	(0.419, 1.104)
	FE	0.473***	0.160	<0.001	(0.160, 0.788)
	RE	0.499***	0.120	<0.001	(0.264, 0.735)
	BE	0.499	0.163	0.002	(0.179, 0.819)
Kowles	OLS	0.167	0.228	0.465	(-0.28, 0.615)
	Cluster	0.167	0.318	0.603	(-0.483, 0.817)
	FE	0.451	0.290	0.120	(-0.117, 1.018)
	RE	0.419	0.352	0.234	(-0.270, 1.108)
	BE	0.419	0.461	0.364	(-0.484, 1.322)
Modified	OLS	-1.682	12.798	0.895	(-26.778, 23.415)
	Cluster	-1.682	17.772	0.925	(-38.031, 34.667)
	FE	0.520***	0.079	<0.001	(0.365, 0.675)
	RE	0.532***	0.448	<0.001	(0.386, 0.678)
	BE	0.532***	0.101	<0.001	(0.334, 0.729)
		Legend:	p < 0.1 = *	p < 0.05 = **	p < 0.01 = ***

CHAPTER SIX: DISCUSSION

In summation, this thesis attempts to answer the question of how MLB teams attempt to maximize their profits. Profits and revenue are largely driven by game attendance; therefore, the paper focuses on what influences attendance to MLB team's home games. This chapter will summarize the major points in the paper, quantify the statistical analysis, and draw the final conclusions based on the results. We should care about attendance in major league baseball because MLB has profit figures in billions of dollars. This provides a large market, and amount of money that could be left on the table if the teams in that market do not take every opportunity to maximize their profits. Teams make almost all of their money from attendance, TV deals, and product sponsors. Of the three, TV deals and sponsors are largely driven by attendance at the games and viewership on TV. While a few papers in this literature look at TV viewership, it can be assumed that if an individual is willing to attend the games, they can just as easily watch them on TV. This high correlation implies that attendance is a driving figure for how teams earn their revenue.

Different models and statistical tools to determine different ways for teams to maximize their attendance are analyzed in this paper. The primary goal is to test if the relationship studied by Knowles et al. in 1992 still holds up in 2013. The replicated model is then modified with the introduction of panel data and additional independent variables. Each of these changes made was meant to improve the quality of the study and determine whether the relationship remains valid, and if there are other variables that could have a strong effect on attendance. The uncertainty of outcome hypothesis states that one of the primary drivers of attendance is that either team has a chance to win the games. Competitive balance is the main factor is determining if either team playing will have a chance to win the game. Competitive balance in baseball is what makes each game interesting and causes fans to attend games. Knowles et al. took this hypothesis a step farther by determining that the optimal expected winning probability of the home team that maximized attendance was 0.6.

The statistical findings in this paper replicated Knowles et al. original model, but refreshed the data for the 2013 MLB season. Knowles' et al. findings were replicated because of structural changes to MLB including expansion teams, relocation of teams, and interleague play. The revised model tests if the relationship between the probability of the home team winning and attendance exists. The replication is important because it also provides a baseline model to judge if the improvements in the modified model are helping, and to see if the relationship between the uncertainty of outcome hypothesis and attendance still hold up.

The betting odds, or money lines, for MLB games were used to calculate subjective probability that the home team would win. A number of different graphs, statistical models, and equations were generated from the data. Panel data techniques to control for variation between teams and across the 81 home games each team played were among the various models generated. Every model, except the naïve model, failed to reveal a statistically significant relationship existed between the probability of the home team winning and attendance at a baseball game. The optimal subjective probabilities that maximized attendance did test significant at the 0.01 level. A relationship between attendance and the uncertainty of outcome hypothesis exist, and attendance is maximized when the home team as a 0.52 chance of winning any home game.

The optimal subjective probability that maximized attendance has a large variance and the confidence interval is wide, which means that the optimal winning percentage for each team could fall anywhere in the range of 0.31 and 0.74. That points to although there is a relationship, it isn't very strong or used commonly because of the large amount of variation and chance to miss the optimal winning percentage. It is not likely that owners and potential consumers use this relationship exclusively. There are many other variables that also tested significant in the modified model that have an impact on attendance as well, such as interleague play, how many weeks into the season the game is, if the opponent is considered a rival, and if the game is expected to be high scoring all tested significant as well and also can improve attendance at games. The primary conclusion from these findings is that an uncertain outcome that slightly favors the home team does not guarantee attendance, but the quality of the game, and opponent matter and have strong impacts on how filled the stands are.

REFERENCES

Berkowitz, Jason P., Craig A. Depken, and Dennis P. Wilson. "When going in circles is going backward: Outcome uncertainty in NASCAR." Journal of Sports Economics Volume 12.3 (2011): 253-283.

Brown, Maury. "Major League Baseball Sees Record Revenues Exceed $8 Billion For 2013." Forbes.com. 17 Dec. 2013. Web. 18 Jan. 2015.

Buraimo, Babatunde, and Rob Simmons. "A Tale of Two Audiences: Spectators, Television Viewers and Outcome Uncertainty in Spanish Football." Journal of Economics and Business. 61.4 (2009): 326-338.

Buraimo, Babatunde, and Rob Simmons. "Do Sports Fans Really Value Uncertainty of Outcome? Evidence from the English Premier League." International Journal of Sport Finance, 3.3 (2008): 146-155.

Butler, Michael R. "Interleague Play and Baseball Attendance." Journal of Sports Economics. 3.4 (2002): 320-334. Print.

Forrest, David, and Robert Simmons. "Outcome Uncertainty and Attendance Demand in Sport: the Case of English Soccer." Journal of the Royal Statistical Society: Series D (the Statistician). 51.2 (2002):229-241.

Horowitz, Ira. "If You Play Well They Will Come—and Vice Versa: Bidirectional Causality in Major-League Baseball." Managerial and Decision Economics. 28.2 (2007): 93-105.

Knowles, Glenn, Keith Sherony, and Mike Haupert. "The Demand for Major League Baseball: a Test of the Uncertainty of Outcome Hypothesis." The American Economist (1992): 72-80.

Rivers, D H, and T D. DeSchriver. "Star Players, Payroll Distribution and Major League Baseball Attendance." Sport Marketing Quarterly. 11 (2002): 164-173.

Schmidt, Martin B, and David J. Berri. "Competitive Balance and Attendance: the Case of Major League Baseball." Journal of Sports Economics. 2.2 (2001): 145-167.

Szymanski, Stefan. “The Economic Design of Sporting Contests.” Journal of Economic Literature 41.4 (Dec., 2003): 1137-1187.

Walker, Bruce. "The Demand for Professional League Football and the Success of Football League Teams: Some City Size Effects." Urban Studies. 23.3 (1986): 209-219.

Whitney, James D. "Winning Games Versus Winning Championships: the Economics of Fan Interest and Team Performance." Economic Inquiry. 26.4 (1988): 703-724.