2003

Effect of competitive balance on single game attendance in major league baseball

Thomas V. Richardson
Colby College

Follow this and additional works at: https://digitalcommons.colby.edu/honorstheses

Part of the Economics Commons

Colby College theses are protected by copyright. They may be viewed or downloaded from this site for the purposes of research and scholarship. Reproduction or distribution for commercial purposes is prohibited without written permission of the author.

Recommended Citation
https://digitalcommons.colby.edu/honorstheses/208

This Honors Thesis (Open Access) is brought to you for free and open access by the Student Research at Digital Commons @ Colby. It has been accepted for inclusion in Honors Theses by an authorized administrator of Digital Commons @ Colby. For more information, please contact mfkelly@colby.edu.
The Effect of Competitive Balance on Single Game Attendance in Major League Baseball

By
Thomas V. Richardson

May 15, 2003
EC 484
Colby College
Honors Thesis
Advisor: Prof. Randy Nelson
2nd Reader: Prof. James Meehan
# Table of Contents

<table>
<thead>
<tr>
<th>Chapter:</th>
<th>Page Number:</th>
</tr>
</thead>
<tbody>
<tr>
<td>Introduction</td>
<td>2</td>
</tr>
<tr>
<td>Literature Review</td>
<td>7</td>
</tr>
<tr>
<td>Variable Descriptions</td>
<td>26</td>
</tr>
<tr>
<td>Explanation of the Model</td>
<td>43</td>
</tr>
<tr>
<td>Empirical Results</td>
<td>46</td>
</tr>
<tr>
<td>Conclusion</td>
<td>60</td>
</tr>
<tr>
<td>References</td>
<td>64</td>
</tr>
<tr>
<td>Appendix</td>
<td>68</td>
</tr>
</tbody>
</table>
INTRODUCTION

Many studies have looked at "competitive balance", but what exactly is it? After surveying much of the literature on competitive balance one finds that there is no clear definition for it. Palomino and Rigotti define competitive balance as: "Competitive balance is measured by uncertainty of the outcome. Fans enjoy more sporting events whose winners are not easy to predict. In other words, the more symmetric the winning chances of the competitors the more exciting the tournament is to watch" (p. 1). However, this is not by any means the definition that all studies use to measure competitive balance in sports. Each study often has a different measure as determined by the author in order to capture what the study is looking for. However, the link between definitions is that they all either try to show how large of a discrepancy in quality exists between the teams of a league, or to capture the uncertainty of outcome of a game.

In recent years there has been much commentary on the health of Major League Baseball (MLB) as a whole. The majority of the cited problems have stemmed from the belief that competitive balance is un-balanced in today's game. The problem with these arguments is that there has been a lot of talk about the issue, but not a whole lot of action in terms of really looking at the problem. Competitive (im)balance became an issue during the 1990's and coincided with the dramatic increase in player's salaries that happened during this time period. The market has become such that only a few teams have the resources to construct a quality team that can compete at an elite level due to the salary constrictions that most teams have. After the 1999 season MLB put out the Report of the Independent Members of the commissioner's Blue Ribbon Panel on Baseball
Economics, exploring the existence of what they defined as competitive balance. This report defined competitive balance as follows: an optimal amount of competitive balance would provide the majority of major league teams the chance at making the playoffs each year. This is based upon the assumption that in baseball today, that while every team is eligible to qualify for the postseason each year, that a large percentage of teams in the league know that they have no chance at making the playoffs before the season even started. This is due to the fact that these teams know that the talent on their team would not allow them to beat perennially more talented teams. This problem was linked to large gaps in the earning potentials of the different teams “markets,” their home cities they play in. The size of a market, in terms of the population of the home cities, is largely correlated to the amount of “local” revenue that a team can raise. Local revenues include money taken in from ticket sales, concessions, memorabilia sold at the park, and parking and other revenue streams generated at the stadium. This is of particular importance because the payroll of a team is largely dependent upon the amount of local revenue that a team can raise. The more local revenue a team raises, the more they can pay for talented players, leading to better teams. The report showed that the vast majority of postseason games played between the 1995 and 1999 MLB seasons were dominated by teams that were in the top ¼ of the league in terms of payrolls. While this study would seem to make sense, it never looked at the effect of competitive balance on individual games, just on postseason success.

The Blue Ribbon study causes one to believe that if a majority of teams don’t have hope for a World Series win at the beginning of the year, then there is little competitive balance. How does this work? Can’t a game still be interesting even though
neither team has a realistic shot at making the playoffs? Can't a team be relatively competitive without qualifying for the playoffs? Also, the past seasons have helped to dispel some of the validity of this study. The 2002 playoffs were full of teams that were not in the top ¼ of the league payrolls. Oakland and Minnesota both qualified for the postseason with payrolls in the bottom ¼ of the league. The World Champion Anaheim Angels had a payroll right around the league median. These results bring up the question of whether the situation that was cited in the Blue Ribbon Panel is a long run problem, or was the time period that was looked at an anomaly? Time will tell. My study doesn't look at payroll issues in particular, but looks at how competitive balance affects the demand for each single game, as measured by the attendance at that game.

Many studies have looked at competitive balance in sports. A majority of the competitive balance studies have looked at panel data sets covering multiple years, however. Many of these studies have looked at the distribution of winning percentages within a league over multiple seasons, looking to see if different teams have an ability to rise to the top of the league, or if the same teams always win (Balfour and Porter 1991; Butler 1995; Quirk and Fort 1997). Schmidt and Berri (2001) use the Gini coefficient to measure competitive balance for baseball seasons since 1900. This measure of competitive balance, the Gini coefficient, measures the amount of inequality existent within a league. It is normally used to measure the amount of income inequality in a population. Other studies have also looked at competitive balance, but they have all looked at competitive balance on a season-by-season analysis. No study to date has looked at how competitive balance affects any part of a single game analysis.
So, the question before beginning this study was how does one measure competitive balance for a single game. Since no study had specifically looked at how competitive balance affects single game attendance, a new measure had to be constructed. Competitive Balance is rooted in how easy it is to predict the outcome of a contest before it is played. For a single game this can be called the uncertainty of outcome of the game.

One way of looking at uncertainty of outcome is to look at the betting lines of the games before they were played. A number of studies have looked at the Uncertainty of Outcome Hypothesis (UOH) by looking at the betting lines of the games. The first study looked at betting lines as a proxy for uncertainty to determine the attendance at soccer matches in Scotland (Jennett 1984), this was followed up by a look at how betting lines in English soccer affected attendance (Peel and Thomas 1988). Borland then looked at the determinants on attendance in Australian Rules Football 1950-1986 (1987), taking into account betting lines as Jennett had. The most recent study looked at single game attendance in baseball during the 1988 season (Knowles, Sherony and Haupert 1992. Clearly uncertainty of outcome has an affect on attendance, but the question is how much?

This study looks at just that; how does the quality of a team affect each single game? To answer this question, three full MLB seasons were used as a data set. A total of 7,279 games were used from the 2000, 2001, and 2002 seasons. By looking at how competitive balance affects the attendance at each of these games, a definitive answer should present itself. In order to look at how competitive balance affects attendance, an econometric model must be created with the single game attendance set as the dependent
variable. The independent variables of the model are different factors that may affect whether or not a fan attends a game.

Many previous studies have looked at the different factors on attendance at sporting events, looking at both seasonal attendance and single game attendance. A large number of these studies have looked at Major League Baseball. Baseball provides a great environment to run studies on due to the fact that meticulous statistics are kept and measures of quality are easily available. Data is both accessible and consistent between all thirty teams.

For these reasons, many studies have looked at the determinants of attendance in baseball. These studies, particularly the single game studies, were instrumental in determining variables to use, and also provided great resources from which to base theory upon. These studies on attendance, however, for the most part do not take competitive balance into effect as a determinant of attendance. Therefore, a new measure of competitive balance had to be created. After reviewing the pertinent literature, a variable measuring for competitive balance for each single game was created. The key variable that was created measures the difference between the winning percentages of the two teams playing, calculated by subtracting the visiting teams winning percentage from the home teams winning percentage (home winning % - visiting winning %). This measure of competitive balance gives the uncertainty of outcome because the greater difference between the two records, the greater amount of certainty that the team with the better record will win. The supporters of competitive balance suggest that attendance will be maximized when there is much uncertainty as to who will win: this happens when the two teams have the same record. When the two teams have the same record the variable
measuring the difference in winning percentages will be equal to zero. The supporters of competitive balance believe that negative effects of this variable on attendance are the same whether the home team had a .600 winning percentage and the visiting team had a .400 winning percentage, or if the home team had was .400 and the visiting team was .600. The reason for doing this study was due to the fact that I didn’t think both examples would draw the same attendance. Therefore I wanted to prove the competitive balance theory to be false.

Growing up in Minnesota, I was always a Minnesota Twins fan. During my prime years to go to the ballpark, between the ages of 12 and 18, the Twins were quite bad. I attended numerous games and observed that the attendance when a team like the Yankees (a perennial powerhouse) came to town was always better than when a team of equal playing ability like the Detroit Tigers came to town. This held even though the uncertainty of outcome would have been much greater for the Tigers game than the Yankees game. This leads me to believe that the competitive balance theory doesn’t hold to be true in MLB. In order to look at the true effects of this variable on attendance, all the other effects on attendance had to be controlled for. Hopefully the results that follow will show that this hypothesis is true. In subsequent sections I review studies that look at the demand for sporting events, explain the model and results of the study, and finally speculate on what these results mean to baseball.

LITERATURE REVIEW
The literature that has been written on attendance in professional sports can be broken into two different categories; you can look at both total season attendance over time, the “macro” approach, or you can explore single game attendance, the “micro” approach. This paper will specifically talk about the micro levels, but much was taken from the macro papers as well.

One of the studies that most closely resembles the research discussed in this paper is “The Determinants of Football Match Attendance Revisited: Empirical Evidence From the Spanish Football League,” by Jaume García and Plácido Rodríguez (2001). In this paper the attendance of football (soccer) matches is looked at for one of the most successful soccer leagues in the world, the Spanish Primera Liga. The paper sets out to create a reliable model to estimate attendance at Spanish soccer games. In doing this they also wanted to estimate the price elasticites for the different variables.

This paper uses data from the four seasons between the 1992-93 through the 1995-96 seasons, which resulted in the use of 1580 observations (games). The Spanish soccer season coincides with the other major European leagues, which take place from September through mid June. The paper looks to determine the affect that different variables have on the attendance of games.

The model that they used includes many different variables. In order to look at the effects on attendance, they set it as the endogenous variable. Also, attendance as the dependent variable is always in log form for all their equations. To account for attendance the number used is the number of total tickets sold for the match, “not including those for children and season tickets” (p. 3). For the independent variables they
used “economic variables, variables proxying the expected quality of the match, those measuring the uncertainty of the result, and those capturing the opportunity cost of attending a match” (p. 2).

The economic variables used included the price of tickets that was calculated by using the least expensive ticket sold, and then deflating it by the CPI. Other key variables used include the per capita income level of the home team province and the population of the home teams province, “which is distributed, when there are two or more teams in a province, according to the number of season ticket holders corresponding to each team” (p. 3).

Variables controlling for the quality of the match included the number of players who had played for their respective national teams, included as a sort of All-Star variable, for both the home and away teams. The payroll of each team is included to give an idea of the quality of each team; this is based upon the fact that greater talent costs more, so the greater the payroll, the better the team should be. A dummy variable is added when the away team is either Real Madrid or Barcelona, as these are historically the two most excellent teams in Spain and therefore may have an affect on attendance when a team hosts them.

Curiously the win/loss record of neither the home team nor the away team is used as a variable. To control for the quality of each team more unconventional variables were used. Included was the home teams record the for the last three games, the score of the last game the home team played, the home team’s position in the season’s standings, and also the chance the home team has of winning the league championship.

Fans enjoy a game where the result is relatively unknown before the game. The
authors of this paper used uncertainty variables for this. Those included the difference in league standings between the two teams. Also used was a variable that measured a team's chance of winning the championship, calculated by taking "the product of the number of games left before the championship is decided and the number of points the team trails behind the leader" (p. 4).

To look at the attractiveness of attending a game, dummy variables were used measuring the weather. The variables were broken into three different categories: "no rain, high temperature; no rain, low temperature; rain" (p. 5). The day of the week the game was played was also determined to have an affect on attendance; they broke this up into weekday games and weekend games. Also used was whether or not the game was televised. Finally, the last variable used was the distance between the two clubs home cities, to control for the possibility of away team fans at the games.

The estimation of the model was done with the Ordinary Least Squares Method (OLS). In order to "estimate price elasticities consistently" (p. 8) the authors chose to use the log form for all of the explanatory variables. The authors found that the income elasticity obtained from the model was "basically positive and therefore attendance is a normal good" (p. 10). These results confirmed the hypothesis that the authors had come to in regards to the affect the variables, positive or negative, would have on attendance. The variables also all turned out to be significant at the 5% level. When the model was switched to a form that did not include logs, the model still performed well, with only a slightly lower R² value.

In general the research found that the group of variables that most affected attendance were the variables that tried to capture the quality of the match. These were
the variables like the national team appearances for each team, and the payrolls of each
team. The second most important variables were the opportunity cost variables, like the
weather variables and if the game would be televised. The authors concluded that the
home teams quality was not more important than the away teams quality, but that in fact
"the impact of the away team quality is clearly higher (on attendance) than that of the
home team" (p. 17). The authors conclude that their results do not support the conditions
that Késenne put forth in his paper "Revenue Sharing and Competitive Balance in
Professional Team Sports" (2000). In this paper Késenne found that the absolute quality
of the game is the most important component in determining revenues. He argues that
this shows that revenue sharing will increase competitive balance and maximize
revenues. However, the results that García and Rodríguez came up with do not support
these results. Since their data showed that the away team's quality has a greater impact
on attendance, "the necessary conditions for revenue sharing having an effect on
competitive balance do not seem to be satisfied" (p. 17).

"Rebuilding Attendance in Major League Baseball: The Demand for Individual
Games," by Thomas H. Bruggink and James W. Eaton looks at the full 1993 baseball
season, the season directly before the strike shortened 1994 season, and tries to determine
the factors that affect attendance on a game to game basis. This is one of four previous
papers that look at individual game attendance in baseball.

This paper looks at which variables affected single game attendance prior to the
strike in 1994 that cost Major League Baseball the second half of that season and the
playoffs. Many fans were disillusioned by the strike and had decided to stop coming to
games. The authors of this paper wanted to determine which factors induced fan interest the most, in order to make recommendations to the owners of baseball regarding gaining back these fans. This was brought on by the clear drop in attendance seen across the board for teams in the 1995 season, after a new labor agreement was reached.

The variables used in the model were chosen for their ability to change fan interest in attending a game, by either enhancing or detracting from the appeal that each game has. The authors look at the baseball games as a good, and "standard microeconomic theory postulates that the quantity of a good or service demanded is a function of its own price, prices of substitute and complementary goods, size of the market, income of consumers, and specific attributes and qualities of the product" (p. 11). So, the good demanded is baseball games, with the quantity demanded measured by the attendance at them.

Attendance at a game is obviously dependent on many things; for this paper the authors created 41 different independent variables to model attendance. They grouped these "explanatory variables into the broad categories of location, expected quality, time and weather, and special factors" (p. 11).

For location variables they were trying to account for the differences in the cities that MLB teams play in. These variables included "metropolitan population, per capita income, average ticket price, the Fan Cost index (a measure of the prices of complementary goods), the number of substitutes (defined as another MLB team playing in the same metropolitan area), the age of the team's stadium, and the percentages of the metropolitan population that are black or Hispanic" (p. 11).

For the expected quality variables, they try to control for the quality of both the
home and the away team. While most of the variables for this category varied, two remained fixed for the whole season. These were the number of division titles a team had won in the three seasons prior to 1993, and also number of games behind the division leader they finished in 1992. The variable factors included “the number of games the home and visiting teams are behind the division leader before each game; the home team’s wins and runs scored over its last ten games; the race, career wins minus losses, and current season wins minus losses of the two starting pitchers; and the number of league offensive leaders and All-Star players in the game” (p. 14). The model did not include any type of variable concerning the winning percentages for either the home or away teams.

The time and weather variables and special factor variables helped to explain attendance differences that were independent of the quality of the two teams, or the city that the game was being played in. The time and weather variables included things such as the time of day and day of the week the game took place, the season of the year, and whether inclement weather was present. The special factors were variables that controlled for possible outliers in the data set. These were things like if there was a promotion going on, if the game was televised locally, if the game was a double header, and if the game was between teams in the same division. Also included were variables that were unique just to the '93 season like a dummy variable that indicated if Nolan Ryan was pitching in the game. His pitching would likely increase attendance as he had announced before the season that '93 would be his last.

These variables were used in a logarithmic model to determine how they affected attendance. The factors for the most part had the signs that the authors expected. Most
of these variables were statistically significant, but some failed to be significant at high levels. For the location variables, both the city’s population and the percentage of black residents in the city were significant and positive. The finding that attendance was higher, ceteris paribus, in cities that had larger black population was contrary to earlier papers looking at how attendance varied with the racial makeup of a city (Noll 1974). However, the variable concerning the Hispanic population of a city was significant and negative. Also, the authors found that newer stadiums result in greater attendance, and that having another MLB team in your metropolitan area decreases attendance, with both variables being found statistically significant.

For the income and price variables, the authors found some interesting results. All the variables were found to be significant, but none acted as they were expected to. The per capita income had a negative impact on attendance, and the price of the tickets and cost of complementary goods (FCI) had positive impacts on attendance. That would mean that “richer” cities had lower attendance, and that as prices on tickets, merchandise and concessions at the ballpark increase, so does attendance, ceteris paribus. “Noll (1974) obtained similar results for per capita income and argued that baseball is an inferior good” (p. 22). In his paper Noll found the negative correlation between income and attendance. He reasoned that “baseball appears to be a working class sport” (Noll 1974, p. 120). He reasoned that baseball tickets were significantly cheaper than tickets to other professional sports (substitutes), and that the slower style of baseball may have been more attractive to working class people whose jobs were very physical in nature.

For the expected quality of the game, Bruggink and Eaton found that the lagged variables for team quality ('92 games back in the division, and division championships
since '90) were both positive and significant. The variables measuring a team's performance during the current year were constrained to the number of games back each team was of the leader of their division, and each team's number of victories over the past ten games, all of which were found to be negative and statistically significant. "In the AL, a home team ten games out of first place playing another team ten games out would have 19 to 26% lower attendance than if both teams were division leaders. In the NL, attendance would be 12-31% lower. If the home team is hot, for example winning eight of its last ten games, it will draw an additional 7-17% in the AL and 13-24% in the NL" (p. 23).

Surprisingly, the number of runs scored in the past ten games has a significant and negative effect on attendance for both the NL and the AL. Fans do not seem to care about the starting pitcher's race, as evidenced by the insignificance of the variable. Also, while the variables for the starting pitchers career and season win-loss records are for the most part significant, they only have a small effect on attendance. The number of All-Stars in the game is significant, however the number of All-Stars on the visiting team has a larger effect on attendance, and the number of home All-Stars was found to have negative effect on attendance as it increased.

Weekend games were found to have higher attendance than weekday games, and games that took place in April and May, before schools get out for summer, attract less fans; poor weather also decreases attendance. Promotions on the day of the game increase attendance, and televised games decrease it by giving fans a free alternative to see the game. Divisional games where the teams playing have more at stake tend to draw bigger crowds. Not surprisingly, Nolan Ryan had a huge positive effect when he pitched
during the '93 season. Attendance could be expected to jump as much as 40% on the
days he pitched.

The authors conclude their study by offering advice to the baseball community in
how to increase attendance. They applauded baseball's decision to increase from four
divisions to six in 1995, although they thought that baseball should alter the schedule to
make it such that teams play divisional foes more frequently. After this paper was
written, baseball did just that, and currently teams do play an unbalanced schedule where
they face divisional foes more frequently. The authors argue for a more balanced league
financially. "Financial viability of each team does not guarantee competitive balance
within a league, but it may be a necessary condition for it" (p. 27). They argue that more
competitive small-market teams will not only gain more attendance at home, but they will
increase attendance on the road, helping the large market teams as well. They reason that
promotions increase attendance, but that marginal benefit analysis should be used to
determine if a promotion should be used. The marginal benefit approach should also be
used when negotiating local TV contracts. The negative impact of runs scored on
attendance should cause the use of the DH in the AL to be looked at closer. The DH was
implemented to increase attendance, but now could be decreasing it.

"The Demand for Major League Baseball: A Test of the Uncertainty of Outcome
Hypothesis," by Glenn Knowles, Keith Sherony and Mike Haupert looks at how the
attendance at baseball games is dependent upon the uncertainty of the outcome of the
game. The authors use the set of National League games during the 1988 season as their
data set.
The paper relies on the uncertainty of outcome hypothesis (UOH) as the grounds for its analysis. "The UOH is predicated on the assumption that fans receive more utility from observing contests with an unpredictable outcome, and posits that the more evenly team playing abilities are matched the less certain the game's outcome and the greater the game's attendance will be" (p. 72). The paper sets out to look at this relationship and also to determine the amount of uncertainty that maximizes attendance.

To test for the relationship between uncertainty and attendance, first a measure of uncertainty had to be determined. The authors created a variable for this by using the betting lines of the games on the day they were played. The authors used the "Eastern line" as their preferred betting line. Using this line the authors were able to create a rather ingenious way to calculate the probability of a team's chances of success (winning). Since the betting line is set as to make each side a fair bet, and to keep the bookie from exposing himself too much, this "expectancy" variable seems to do a good job of conveying the probability of winning.

On top of this key variable, the authors also include other key determinants of attendance. To control for the quality of the game a variable summing the number of games behind in the standings the home and away teams are is used. Dummy variables are added if the game took place on either the weekend or at night. The MSA population is included for the home city for reasons already explained. The per capita income for the city is used to proxy for the economic condition of the area as it has been in previous studies. Adding to the measure of a city's success is the unemployment rate of the city. The last variable used is the distance between the two team's hometowns, so as to account for visiting teams fans traveling to the games and to account to possible
geographic rivalries.

The specification of their model was linear and OLS was used to estimate the model. The authors included both “expectancy” and expectancy-squared variables in the model, with the latter variable included to look at how attendance is affected at the extremes of the expectancy; it was also significant. The two variables “have the correct sign for attendance to be a concave function (of expectancy). This indicates that attendance increases at a diminishing rate with the uncertainty of the game’s outcome up to some maximum” (p. 76).

Neither the unemployment nor per capita income variables were statistically significant. All the other variables had statistically significant values and signs that the authors expected. They found that the variables that had the largest impact on attendance were whether the game was played on the weekend or at night, the population of the city, and the combined games back of the two teams. The distance between the two teams was significantly negative, which could be attributed to the fact that they failed to control for the fact that some cities have two teams, implying that the distance between them is zero.

Using the “expectancy” variable the authors were able to determine what level of certainty maximized attendance. They found that the home team having a .6 expectancy of winning maximized the attendance. This was attributed to the fact that home fans account for the majority of those in attendance, and while they want to see a game with some uncertainty, they would prefer to see the home team win. These results were consistent with what Quirk and El Hodiri hypothesized in 1974 (p. 35).

The results of this paper support the hypothesis that competitive balance is valuable to baseball, and that it would be in MLB’s best interest to support measures to
enhance competitive balance. Attendance has shown to be maximized when the two
teams both have a chance at winning. All teams will benefit if attendance can be
maximized, due to the increased local revenues that are generated with higher attendance.
The authors also raise the point that since it is in the best interest of all the teams if the
home team is slightly favored, that ballparks should be "characterized by idiosyncrasies
around which teams can be built," (p. 78) as opposed to ballparks with standardized
dimensions that don't offer the home team a unique advantage.

In "Factors Affecting Attendance of Major League Baseball: II. A Within-Season
Analysis," by John P. Marcum and Theodore N. Greenstein, the authors looked at game-by-game attendance for two Major League Baseball teams during the 1982 season. The
two teams used in the study were the St. Louis Cardinals and the Texas Rangers, with
each team being chosen for specific reasons. These two teams were chosen in tandem for
the purpose of comparing two very different teams that also had many things in common
to add to the significance of the comparison.

"The Cardinals and the Rangers had very different seasons in 1982... The
Cardinals won their division by three games and defeated Milwaukee in seven games to
win the World Series, while Texas lost 98 games and finished 29 games out of first place.
Perhaps even more important in an analysis of attendance, the Cardinals were in first
place for over three-fourths of their home openings in a hotly contested race whereas the
Rangers were never competitive" (p. 316).

Even though these two teams had a vastly different product on the field, the two
clubs shared many similarities. The population of the two metropolitan areas the teams
played in were within 150,000 of each other, making it possible to leave the population variable out of the model. Also, the two teams had similar Television deals, nullifying the effect of different deals would have on attendance. These teams also represented different leagues, St. Louis in the National League and Texas in the American League.

The data set used a total of 158 games, 78 for Texas and 80 for St. Louis. The authors used a total of ten explanatory variables in their model: “day of the week, type of team promotion, opposing team, opponent’s won-lost percentage, opponent’s games behind first place, home team’s won-lost percentage, home team’s games behind first place, home team’s record over the previous 10 decisions, weather conditions, and whether the opening involved a doubleheader” (p. 317). Dummy variables were used for team promotions and the weather variable. Team promotions were broken up between major promotions, like bat night or cap night, and minor promotions. To measure for the presence of bad weather, three dummy variables were classified as either a day with rain, a temperature 5 degrees less than the average in April and September, or a temperature greater than 5 degrees warmer than the average in the other months (p. 317).

The dependent variable was attendance, however it would be hard to compare attendance since Busch Stadium in St. Louis could sit almost 9,000 more people than Arlington Stadium. To account for this the authors used attendance divided by capacity, or the percentage of the stadium filled, as the dependent variable.

The results showed that, on average, St. Louis filled .526 of its stadium with Texas filling .358. This is even greater considering the difference in capacities of the two stadiums. Busch Stadium, where the Cardinals play seats 50,222, whereas Arlington Stadium, where the Rangers play, only seats 41,284. Therefore the Cardinals averaged
26,417 fans and the Rangers averaged only 14,870 fans per contest. This accounts for a 79% difference in true attendance (as measured by fans in attendance), but is made to look smaller with the "percentage filled" dependent variable. Since this dependent variable is proportion of seats filled in the stadium, the independent variables "are interpretable as the increase in proportion of seats filled attributable to a specific factor" (p. 317).

The results of the test were somewhat inconclusive; "only two of the factors (in addition to the intercept value) are significantly different from zero for both ballclubs in the various models: Saturday attendance and major promotions" (p. 318). This is alarming because some variables are statistically significant for one team, but not the other. This may lead one to believe that this model isn't appropriate for both of the clubs. However both regressions had very high R² values, with St. Louis at .66 and Texas with .88.

The authors explain the differences are most likely due to the contrasting quality of play the home team displays. For example, "when a home team is doing well, attendance will be good regardless of which club is in town or what the management is giving away at the ballpark. When a team is doing poorly, however, fans need an extra incentive to show up; thus, giveaways and first-place opponents increase attendance" (p. 319). This was demonstrated in the results in that the Rangers attendance was much more dependent upon the opponent they were playing and the existence of a major promotion. The authors found that the opponent accounted for 28.7% of the variation in attendance for the Rangers, but only 14.3% for the Cardinals. The conclusion that the authors were able to make between both of the clubs was that "most of the variation in
attendance can be attributed to three factors: opponent, day of the week, and promotions" (p. 319).

The competitiveness variables for this model, home and away winning percentages, and home and away games behind first place, were very hard to interpret. The values that are given for these variables in a table of regression results within the paper do not coincide with the conclusions that the authors come up with when explaining the results of their research. In the paper the authors state that the quality of an opponent has a positive effect on attendance, however, all the variables in the table had negative values, indicating a negative effect on attendance. Also, the paper focused very little on the interpretation of the variables, leaving the reader to glean as much as they could from the table that was included.

This paper explored in detail the effect of different variables on the single game attendance of two different teams in MLB. The authors noted to use caution in using their results in extrapolating results for other teams (p. 320) since there are major differences between the two teams used other than winning percentages. They go on to add that socioeconomic factors such as per capita income may be types of variables that should be looked at in future studies.

Many of the studies that have been done regarding attendance in baseball have used cross-sectional data using periods of years as their data sets. These studies have looked at attendance in baseball at the seasonal level, looking at large periods of time, to determine how attendance to baseball games has historically been determined. While these studies don't look at some of the factors that are important for single game
attendance, the theories and ideas of the authors have greatly influenced the research of
the single game studies on attendance.

The first such study on attendance was by Demmert (1973) in his book, The Economics of Professional Team Sports. This study looked at attendance for 16 major league clubs from 1951-1969. His model included “a combination of cross sectional and time series data” (p. 58), to estimate the equation. This equation included many important variables like average ticket price, the average family buying power in a metropolitan area, with both of these numbers being discounted with the CPI. The SMSA for a city was used for the population variable. To account for team quality during the season, the games behind the divisional leader was taken at 5 different times during each season. Dummy variables were added if a team was a defending league or world champion. To control for substitutes, both direct and indirect, both the number of other professional baseball teams and other professional sports team in the same metropolitan area are included. The last few variables include variables measuring the existence of a pennant race during the season, the number of total home and away games that were televised, the number of years a team had been in a city, whether the stadium the team played in was new, and finally a dummy variable for when MLB increased its schedule from 154 to 162 games and the expansion of 2 teams in each league in the early 60’s.

As was found in the other studies, the population of a team’s metropolitan area has a large effect on the attendance of that teams games. Demmert also found that a close race for the pennant does not seem to greatly affect the demand to the games; what is important to demand is that the team wins this battle. “We must conclude on the basis of
this evidence, therefore, that the incentive of the individual club is to win, and not necessarily to win by a close margin" (p. 67). The existence of substitutes in the metropolitan areas has a negative effect on the attendance of baseball games. Demmert estimates that for a city of 2.5 million the existence of another MLB team in the area decreases season attendance by roughly 250,000, and the existence of another pro sports team other than baseball decreases attendance by 300,000.

Demmert’s research was groundbreaking in the study of professional sports economics; his study was the first to explore the demand of baseball. The studies that have followed have all drawn from this study, in both theory and the use of his variables and model.

The next major study done upon season attendance was by Noll (1974) in his book *Government and the Sports Business*. This paper followed much the same route of Demmert’s work, but added some new factors that may have had a hand in his “new” and, in some cases, different results.

This study looked at the 1970 and 1971 seasons for all the teams in MLB. The same general factors were included that are accounted for in most demand functions, such as metropolitan population, per capita income and ticket price. However he had some interesting results for these variables that Demmert did not have. He found that attendance is negatively related to the per capita income level of a city. He reasoned that baseball may be an inferior good, and that could be explained by the fact that “baseball appears to be a working-class sport” (p. 120). Otherwise his results for these variables was consistent to what Demmert had found for these variables.

Noll also looked at a city’s size as a proxy for the amount of entertainment competition that a professional sports team would have for fans. He reasoned that the
larger a city's population, a greater number and diversity of options for entertaining one are present. For instance, the New York teams must compete with the musicals of Broadway for fans, whereas the Kansas City franchise has no comparable competition. The result is that the size of a city's population will have conflicting effects upon baseball attendance; the sheer size of people to draw from will cause attendance to increase, but the vastness of substitutes for entertainment will cause attendance to decrease. The results of Noll's models supported the hypothesis that population has a positive effect on attendance, however, his reasoning brings up an interesting concept.

Noll's research included variables that were scaled by population. He used the SMSA population to create interaction terms with each variable to account for the effect that population had on each variables coefficient. Due to the difficulty in interpreting these variables Noll then showed the variables effects on attendance for three different sizes of cities. He used hypothetical city population sizes to show the different results for "average" sized cities represented in MLB. The three sizes were: "1.5 million (about the size of the smallest cities having baseball teams), 3.5 million (the baseball-wide average, and roughly the size of the sixth largest SMSA having a baseball team), and 12 million (representing New York)" (p. 120).

Other than population Noll used a variety of variables. He included the average ticket price, stadium age, number of star players ("star" players were selected by the personal judgement of the author), and percentage of black population. Also included were the other professional sports teams competing in the city, to control for competition for fans. To control for the quality of the play on the field Noll used variables like the existence of the team being involved in a close pennant during the season, the last time a
team won the pennant and he included the number of games a team was out of first place at different points during the season.

Noll found that season attendance can be increased greatly with a recent pennant win. The presence of stars can greatly increase attendance, although with a decreasing marginal product for more than one star. The presence of other sports teams decreases season attendance, ceteris paribus. The rest of Noll's results coincide with the results that have already been cited. The question that Noll brings up about baseball being a working class sport has brought up much debate in later papers, and this study is almost always cited when talking about attendance in any sport.

VARIABLE DESCRIPTIONS

Building a model to measure the significance of competitive balance on stadium attendance on a per-game basis necessitates the use of many variables to control for reasons that a fan would, or wouldn't attend a Major League Baseball game. Many of these factors are obvious, while others are not apparent at first.

The first variable to describe is the dependent variable, single game attendance or ATTENDANCE. This variable was obtained, along with many others, through the use of www.ESPN.com. The attendance of Major League Baseball games, since 1994, has been measured by the total number of tickets sold for that game. This variable measures the demand for baseball in that the product is each game. To measure the demand of this product, you only need to look at the consumers of that product, the fans that attend the
game, measured by the total attendance. What differentiates this product is that each different game has nuances than can affect the demand for that single game. This variable is consistent with all past studies as the dependent variable when determining the demand for an athletic competition.

The first and possibly most obvious factor to a team's attendance is the population of the city that the home team plays in, POP. This "market" that the team plays in establishes the core fan base that will come to watch the team at the park on a regular basis. The population of the cities in the United States was found using U.S. Census information from the 2000 census in The Statistical Abstract of the United States: 2001. For the two Canadian cities that have teams, data was collected from www.statcan.ca, where the Metropolitan Statistical Areas for those cities was taken from the 2001 Canadian census. The 2000 population was used for all three years because estimates for 2001 and 2002 were unavailable.

Population was measured using the Metropolitan Statistical Areas or the Primary Metropolitan Statistical Areas and Consolidated Metropolitan Statistical Areas when available. Populations in the statistical abstract are measured in the thousands, so a population of 1,000 actually means 1,000,000 inhabitants. Since the populations of metropolitan areas are measured in 1,000 people increments, a one-unit increase actually represents an increase in 1,000 people. MSAs are cities that have over 50,000 inhabitants, or "census defined urban areas of over 50,000 within a metropolitan area of over 75,000" (www.census.gov). An MSA that has over one million inhabitants is given the designation of CMSAs if "separate component areas can be identified within the entire area by meeting statistical criteria specified in the standards, and local opinion

1 A brief description of each variable used in the final model can be found in the Appendix (A-I)
indicates there is support for the component areas. If recognized, the component areas are designated PMSAs, and the entire area becomes a CMSA. PMSAs, like the CMSAs that contain them, are composed of entire counties, except in New England where they are composed of cities and towns. If no PMSAs are recognized, the entire area is designated as an MSA” (www.census.gov). In simpler terms a CMSA is an amalgam of cities that make up a concentrated population, such as Los Angeles and Orange County, or the greater New York area. Within the CMSA are separate cities, designated the PMSAs, which would just be New York City or Los Angeles, not including neighboring cities within the metropolitan area that makes up the CMSA.

Sports teams create a regional fan base, where many fans who come to the games and root for the teams are not located within the city limits of a team. By using the population of the metropolitan areas, a better model is built by controlling for the number of people that a team pulls from for its core fan base. By using the CMSA, when available, it gives the best idea as to how many people could be prospective fans. However, the MSA of a team is most likely where the greatest number of fans who commute to games would be, creating a valuable number for the model as well. When a city is not large enough to be designated a CMSA, the MSA is an equal to the CMSA in showing the size of the metropolitan areas. For these reasons the CMSA population is used whenever possible, with the MSA population used if no CMSA is available. This variable is called CMSAPOP.

While the size of a market increases the amount of people to draw fans from, creating a larger demand, the size of a city may also have a countering, and negative affect on attendance. As Noll (1974, p. 117) noted, the size of a city is also a proxy for
other entertainment options. As a city becomes larger, the availability of activities other than professional sports becomes more prevalent. For example, the two New York teams have to compete with Broadway productions and a Major Opera House among others for a fan's attention; this is not the case in a smaller market like Kansas City. With that said, it is expected that this variable will still be positively correlated to attendance.

Although the population of a city is very important to determine attendance at games in that city, it is not the only important factor about that city. The per capita income of the city also influences the attendance of games. This variable is called INC. Professional sports are expensive to attend, and therefore the wealth of a city as a whole can greatly influence a team's attendance at games. The data for the personal per capita income of each U.S. city was taken from The Survey of Current Business, which gave the income levels for 2000 that had been calculated by the United States Census Bureau. MSA data was used for all, with CMSA data augmenting the data in cities where it was available. For the Canadian cities, data was again taken from the 2001 Canadian Census at www.statcan.ca for the individual income of each city in Canadian dollars (it was assumed the figures were in Canadian dollars). These figures were then converted to U.S. dollars using the historical exchange rate for January 1, 2001, as quoted from the Bank of Canada website. The 2000 (for U.S. cities) and 2001 (for Canadian cities) per capita income level were used for all three years because estimates for the other years were unavailable. A variable measuring the wealth of a city is commonplace in attendance studies. Knowles, Sherony and Haupert (1992), Borland and Lye (1992), Bruggink and Eaton (1996) and Garcia and Rodriguez (2001) used per capita income in
studies looking at how single game attendance is determined for sporting events. This variable is expected to have a positive coefficient.

Wealth of a city is only significant when one considers how expensive the tickets to the game are. To determine the price of the tickets, the average ticket price for the stadium is used. This data was provided by Team Marketing Report, a company that calculates sports business numbers to enable teams to maximize profits. Each year they calculate the average ticket prices for each team. Also used to determine the price of the games was the Fan Cost Index™, a measure calculated by TMR that includes "two adult average price tickets; two child average price tickets; four small soft drinks; two small beers; four hot dogs; two programs; parking; and two adult-size caps" (www.teammarketing.com). This number helps to give the “true” price of a game for a family, taking into account other costs that may impact the decision to attend a game.

For the model, $FCI$ or the fan cost index is used. This variable captures the price of the game better than just the average ticket price by measuring the price of complementary goods with it (Bruggink and Eaton p. 11). This variable is relatively new, originating in 1992, and therefore could only have been used by the newest studies. Bruggink and Eaton used FCI in their paper looking at repairing attendance in baseball after the players strike in 1994. Also, the fan cost index is highly correlated with ticket price, because ticket price is included as one of the factors in FCI. So, ticket price does not have to be included as a variable, as it is represented by FCI. As $FCI$ is a cost variable it is expected to have a negative sign.

The next explanatory variable is another variable that measures the competition that a Major League Baseball team faces, although not in the traditional sense. A team
competes against the players on another team, but a few teams in MLB face another type of competition. If two teams play within the same CMSA as another, they are in essence competing for fans with another, equally accessible team. To account for this a dummy variable was added to the model. This dummy variable is called COMP, which gives a team a value of 1 if another MLB team plays within its CMSA, and a 0 if not. This variable only applies to eight teams. New York and Chicago both support two MLB teams, the Anaheim Angels share Orange County with the Los Angeles Dodgers, and the San Francisco Giants share the Bay Area with the Oakland Athletics. This variable is used in numerous previous studies (Demmert 1973; Noll 1974; Coffin 1996), to control for competition for fans. Since this variable measures the existence of a direct competitor within the CMSA, its hypothesized sign is negative. One would expect another close substitute within the area that a team draws its fans will cause attendance to decrease, ceteris paribus.

The last of the variables that don’t change during the season have to do with the stadium that the home team plays in. Many previous studies have used the age of the stadium to control for the type of environment that a team plays in. Their reasoning was that a newer stadium would be more “fan friendly” and that a newer stadium would cause attendance to increase. Four previous studies have looked directly at the age of the stadium in regards to its effect on attendance (Demmert 1973; Noll 1974; Bruggink and Eaton 1996; Coffin 1996). However, it doesn’t take a new stadium to draw a large crowd. Just look at the three oldest parks in baseball: Fenway Park (Boston Red Sox), Wrigley Field (Chicago Cubs), and Yankee Stadium (New York Yankees); these “classic” parks don’t have any problem bringing fans in. Actually these classic parks can
help to increase attendance by fans that want to see baseball history firsthand. Since these very old parks can still draw them in, and would be extreme outliers, it was decided to not include the age of the stadium as a variable.

To control for different types of stadiums dummy variables were added. Teams that played their home games in a domed stadium were assigned the variable $DOME$, whereas teams that played in a park with a retractable roof were assigned the variable $RETRACT$. Since the omitted stadium is a conventional, open-air stadium, the coefficients of $DOME$ and $RETRACT$ will be the number of fans +/- difference that a team can expect than if they played in a conventional park.

The information on each stadium was provided by each teams official website, which are all accessible under the umbrella of www.mlb.com, Major League Baseball’s official site. The current popular feeling is that baseball should be played outdoors, that dome stadiums draw fewer people to the ballpark, therefore $DOME$’s hypothesized sign is negative. Retractable roofs on stadiums are a relatively new feature that only new stadiums have. These stadiums are newer, more fan friendly, and can block out bad weather. For that reason $RETRACT$ is expected to possess a positive coefficient.

The attendance of different teams for games can be compared, however attendance is capped at an upper bound by the capacity of a stadium. While there is no way to determine the excess demand of a full stadium, you can account for the size of a stadium when running regressions. Each teams home stadium capacity was included as the variable $STDCAP$. Marcum and Greenstein used stadium capacity in their 1985 study comparing the attendance of St. Louis Cardinals and the Texas Rangers by dividing attendance by capacity to calculate the percentage of the stadium that was filled. As
STDCAP measures the number of people that can physically fit in a ballpark it is expected to have a positive coefficient just due to the fact that more people can fit in a bigger stadium.

The rest of the variables that are explored in the model vary from day to day; capturing what truly effects the demand of a single game. Factors such as the time of day that the game was played, and whether the game was part of a double header were found in previous single game attendance studies to be important variables in explaining the attendance of a game (Marcum and Greenstein 1985; Knowles, Sherony and Haupert 1992; Bruggink and Eaton 1996). For this data the baseball Hall of Fame's research library provided the schedules for each team from various sources; that included both the days and times of each game. If a game was scheduled to begin before 5:00 p.m. local time it was given a value of 1 for a “day game” dummy variable, all games beginning after 5 local time were given a 0 for that variable; this variable is called DAYGAME. If a game was either the first or second leg of a double header it was given a value of 1 for the dummy variable, a 0 if it wasn’t; this variable is called DOUBLEH. These variables help to control for the possibility that some would be fans wouldn’t attend a game during the normal working hours of 8:00 am to 5:00 p.m.

The prevalence of double header games has diminished over time and it is difficult to determine the affect that the variable will have on attendance. For that reason it is difficult to determine the sign of DOUBLEH. Day games are also difficult to determine the affect they have on attendance. A day game over the weekend would most likely not have a great difference in attendance than if the game was held at night. However, when a game is played during the day during the week, different effects can be
observed. For example, the Chicago Cubs traditionally play the majority of their games during the day, never having difficulty filling Wrigley Field. That the Cubs play most of their games during the day is very well known and acts as an added incentive to attend their day games; it has been romanticized. However, for most other teams, they schedule few day games during the week due to poor attendance during the working hours. The opportunity cost for fans is greater during the day due to the lost wages that they incur. For this reason the hypothesized sign for \textit{DAYGAME} is that it will have a negative effect on attendance.

A clear determinant in the demand of a baseball game is the day of the week that the game is played. All previous (Hill, Madura and Zuber 1982; Marcum and Greenstein 1985; Knowles, Sherony and Haupert 1992; Bruggink and Eaton 1996) studies looking at single game attendance in baseball have explored the effect that the day of the week has on attendance. The day of the week that the game was played was obtained through www.espn.com, by accessing each teams "schedule" page within the site. Each day is given its own dummy variable, measuring the effect that each day has on attendance, \textit{ceteris paribus}. In some previous papers a variable was created to account for a midweek game instead of giving each day its own dummy variable (Hill, Madura, and Zuber 1982; Knowles, Sherony and Haupert 1992; Bruggink and Eaton 1996; Garcia and Rodriguez 2001). However, for this study we tried a "midweek" variable and tested to see if it reacted the same way as individual dummy variables for each day did using the Log-Likelihood test, where we found we could not package these variables as a "midweek" dummy. It is expected that games played on Friday and the weekend will have greater
drawing power than games played during the week but the exact hypothesized signs for each day cannot be determined before the regressions are run.

Inclement weather can have a large affect on the attendance of a baseball game. Obviously days that don’t have nice weather are less desirable to attend than nice days, ceteris paribus. Each study done on single game attendance in baseball has used a variable to measure the weather on game day, each with a different way of determining how they wanted to control for weather. To calculate this variable the high temperature, low temperature (measured in Fahrenheit degrees) and precipitation (measured in inches) in the home city for the day of the game was obtained. This data was collected from the National Climatic Data Center, a branch of the National Oceanic and Atmospheric Administration, for both U.S. and Canadian teams via the NCDC’s website. Two variables were then calculated; $AVGTEMP$ was, as it looks like, the average temperature on the day of the game, calculated by taking the average of the high and low temperatures. $PRECIP$ is a linear variable measuring the amount of precipitation the home city experiences during a given day.

However, not all teams experience weather the same way. As shown earlier, some teams play in a climate controlled environment of a dome, others play in the semi-protected confines of a retractable roof stadium, while the majority of teams must face the full fury of mother nature on a daily basis playing in open-air stadiums. To account for this, the teams that play in a domed stadium were given a dummy variable (as explained earlier). These domed teams were given low temps of 70, high temps of 75 and 0 inches of precipitation across the board to account for their climate controlledness. The retractable roof stadiums create a dilemma for the weather variables; due to the fact
that each one has its own distinct design, they each have differing abilities to block out inclement weather. So, including the weather data for these cities was still necessary.

The hypothesized sign for \textit{AVGTEMP} is that greater temperature creates greater attendance. While the people in Phoenix may disagree with this considering it can't be that enjoyable to watch a baseball game in 115° degree heat. However the majority of MLB teams play in more temperate zones that would embrace greater heat. \textit{PREcip} is expected to be negative; rain means rain delays and wet seats.

The next two variables are the two most important variables to this paper. They are the measures of competitive balance this paper relies upon. \textit{WINDIFFP} and \textit{WINDIFFN} measure the same thing, but create vastly different results in the model. As is, the measure of competitive balance is the difference in the winning percentage of the home team and the winning percentage of the away team (home winning percentage - away winning percentage). However, there are different implications on attendance if the home team has a better winning percentage than the away team or if the home team's winning percentage is worse than the visiting team. For all the games, the differences in winning percentages were calculated. If the value was greater than 0, meaning the home team had a better winning percentage, and then that positive value was given to represent \textit{WINDIFFP}. If the value was less than 0, \textit{WINDIFFP} was given a value of zero. If the value of the difference in winning percentages was less than 0, meaning the visiting team had a better winning percentage and then that negative value was given to represent \textit{WINDIFFN}. If the value was greater than 0, \textit{WINDIFFN} was given a value of zero. In
order to keep the variables all positive in order to more clearly be able to see how they affect attendance, $WINDIFFN$ was multiplied by negative 1, giving it a positive sign.\(^2\)

These variables are particularly interesting because of the implications of each variable. There are two different effects each of these variables brings into the equation; one that could support the need for competitive balance, the other discourages it. The argument for competitive balance would assume that attendance would be maximized where there was total uncertainty as to who would win the game, this would be represented with the teams both having identical winning percentages; and therefore when both of these variables were equal to zero. To conceptualize this effect just think of attendance being maximized where the teams have the same record, with attendance dropping roughly equally for the deviation from that. It shouldn’t matter whether the away team is better or worse than the home team, they should both effect attendance negatively. This is due to the fact that having one team being more dominant (with a better winning percentage), there is less uncertainty as to who will win; the team with the better record will be favored to win.

The second effect causes the competitive balance theory to be lessened in that it is expected that attendance increases as the quality of opponent increases. This argument is based upon the hypothesis that fans attend games that feature the elite of that sport; regardless if the home team has a chance to win. This is akin to fans of the Tampa Bay Devil Rays (who had a rather poor three year stretch over the course of this study) attending a game in greater numbers to see the home team play the superior New York Yankees (the current elite team in baseball), even though the Devil Rays have little chance of winning the game, ceteris paribus. However, the flip side to this hypothesis is

\(^2\)A graph helping to visualize the $WIDIFF$ variables can be found in the Appendix (A-2)
that one would expect the home fans of Yankees to more sparsely attend a home game against the poorer Devil Rays even though the Yankees have a very good shot at winning the game.

Based upon these two effects, both would cause $WINDIFFP$ to have a negative sign, negatively affecting attendance. However, for $WINDIFFN$ the two hypothesis have conflicting effects on attendance. Competitive balance would argue that any deviation from the same record would cause attendance to fall. The other theory supports that attendance will increase if a superior team comes to town. It is expected that the competitive balance theory is not as strong as it has been made out to be, and that better teams will cause attendance to increase, ceteris paribus. This means that $WINDIFFN$ is expected to have a positive sign.

A team's record is important to fan interest, but of equal importance is the number of games the team is out of first place. Using the data from espn.com, it was possible to create a spreadsheet to determine how many games each team was out of first place in their respective division during the season; this is the variable $GAMESBACK$. Also calculated was the number of games out of first place the visiting team was on the day of the game, this variable is called $ABACK$. Marcum and Greenstein (1985), Knowles, Sherony, and Haupert (1992), and Bruggink and Eaton (1996) all used some form of the number of games back a team was in its division as a variable in their single game attendance models. The number of games a team is out of first place is a proxy for how successful a team is. As this number becomes larger, it is suspected that it will decrease attendance, ceteris paribus. The coefficients for these variables, therefore, are expected to possess negative signs.
Although the general quality of teams is important in determining attendance, certain players can cause attendance to increase regardless of the quality of their team. To control for this, the number of All-Stars on both the home and away teams was included. Noll (1974) used a variable to account for the number of "star" players the home team had. Bruggink and Eaton (1996) used variables measuring the total number of All-Stars on both the home and away team's rosters. All-Star data was collected from the online Baseball Almanac (www.baseball-almanac.com). Since the all-star game is played at midseason, there is the problem of determining the number of All-Stars a team has on it during a year. To rectify this problem, at the beginning of each season each team is credited for each player they have on their roster who played in the All-Star game the year before. Therefore a team was credited with an All-Star at the beginning of the season whether they retained the services of their own All-Star in the offseason or signed/traded for a reigning All-Star in the offseason. After the All-Star break each team is credited with each All-Star who was selected to the game from their team. These factors are included for the reason that many fans will attend a game to watch their favorite star, with more stars causing attendance to increase. As fans come to watch a visiting teams All-Stars as much as they come to see the home teams All-Stars, each is accounted for. HALLSTARS measures the number of home team All-Stars while AALLSTARS measures the number of away team All-Stars. Both of these variables are expected to positively affect attendance and therefore should have positive signs.

As this study looks at data covering three different seasons, these years have to be controlled for. To do this, dummy variables were created to account for the year. TWOTHSTNDONE is a dummy variable that is given a value of 1 for all the games played
during the 2001 season, with a value of 0 for the other years. *TWOTH SNDTWO* does the same for the 2002 season. The 2000 season is accounted for by the fact that it is the omitted variable from the equation, the other two seasons will show the difference in attendance they have in comparison to the 2000 season.

Baseball is broken up into six different divisions, with three in each the National League and American League, with each division having between four and six teams with the majority (4) containing five teams. To control for this the dummy variable *DIV* gives a value of one if the two teams competing are from the same division. Bruggink and Eaton (1996) included a divisional variable in their study of baseball attendance. In baseball the winner of each division goes on to the playoffs; also, divisions are set up so as to capitalize on regional rivalries. The fact that divisional win-loss records serve as tiebreakers for who gets to continue into the playoffs, these games also have an added value of intrigue to them. These effects all will cause attendance to increase, ceteris paribus. However, the MLB schedules of today put a "weighted" schedule, where teams in the division play upwards of 20 games apiece against one another. This can cause a fan to get tired of watching the home team playing the same opponent over and over. Much of the enjoyment of watching a baseball game includes observing how your hometown team stacks up against other teams, and other players. By playing the same team repeatedly, especially if they are below average, this can have a detrimental effect on attendance. For these reasons the sign of the variable coefficient for *DIV* cannot be determined before the regressions.

A baseball season is a long haul; each team plays 162 games, 81 at home and 81 on the road. This season is also long for the fans. To account for the length of the season
GAMESLEFT was added to the model. This variable measures the number of games remaining in the season by taking 162 and subtracting the sum of the home teams wins and losses (total games played to that point in the season). This variable captures two different, but complementary effects on attendance. In the early months of the season, April and May specifically, the attendance to baseball games historically sags in comparison to the attendance of the summer months (disregarding the home openers that traditionally draw large crowds). As Bruggink and Eaton stated, this can somewhat be attributed to both the fact that these months are noticeably colder than the summer months and schools are still in session (p. 24). They also attribute this to the fact that the NHL and the NBA have their playoffs scheduled during the early months of the baseball season. Also, as the season draws to a close, every game becomes more important in the drive to qualify for the playoffs, causing attendance to increase. These two effects serve to support a sort of crescendo of attendance towards the end of the season. GAMESLEFT starts the season at 162 dropping to 0 on the last day of the season, for this reason it is expected to have a negative sign.

The next three variables are lagged dummy variables measuring the success of a team in the playoffs from the season before. PRVPLYOFF gives the home team a value of 1 for the season if they were able to qualify for the playoffs the season before. PRVWS is given a value of one for the two home teams that qualified for the World Series the year before. PRVWSCHAMP is a variable that gives a team a value of one if they won the World Series the year before. Demmert (1973), Noll (1974), Coffin (1996), and Bruggink and Eaton (1996) all used some sort of lagged variables measuring a teams previous luck in the playoffs. As they all found these type of variables to be positively
associated with attendance, and that logic would support this it is assumed that the coefficients of these variables would be positive.

The last four variables are all interaction variables containing variables that have already been explained. The first two measure the combined effect that the games left in the season have with the differences in the two teams winning percentages. \textit{GLWINDFP} and \textit{GLWINDFN} measure this by multiplying the two variables \textit{GAMESLEFT} and \textit{WINDlFFP} (or \textit{WINDIFFN}). These variables look at how the differences in winning percentages are affected by the time in the season that it is. It is expected that as the season gets nearer to the end, and the winning percentages get closer (by the competitive balance theory), that attendance will increase. These two variables measure the effect that competitive balance has at times during the season; it is difficult to determine the sign of these variables a priori.

The last two variables are interaction terms between the difference in winning percentage (\textit{WINDIFFP}, \textit{WINDIFFN}) variables and the divisional dummy variable (\textit{DIV}). These two variables are calculated by multiplying the two aforementioned variables and are called \textit{DIVWINDIFP} and \textit{DIVWINDIFN}, respectively. Since the divisional dummy variable only has a value when the teams playing are in the same division, these variables will also only have a value when the two teams share a division. These variables will hopefully show the relationship that rivalries have with the difference in winning percentages. Again, it is difficult to determine these variables signs before the regressions are run.$^3$
EXPLANATION OF THE MODEL

When running regressions a functional form of the model must be chosen. For this study both the linear form and the log-linear form were explored. In linear form both the dependent and independent variables are unaltered. "The slope of the relationship between the independent variables and the dependent variable is constant" (Studenmund 2001, p. 202). The log-linear or semilog functional form that was used changed the dependent variable into its natural log, with the independent variables remaining the same.

Each of these models has its benefits and its drawbacks. The linear form is the form most often used in econometric modeling. Since the variables have a constant slope against the dependent variable, the coefficients are very easy to interpret. The semilog form is very useful in that it doesn’t constrain variables to have just a linear relationship with the dependent variable. In addition, the coefficients represent the percentage change in the dependent variable given a one unit change in the independent variable.

The first step was to determine the appropriate functional form for the model. The R^2's may not be compared directly because the dependent variables aren't the same. It is possible, however, to test whether one functional form dominates the other. A statistical test was used to determine which form provides a better fit (Wooldridge 2000, pp. 202-204).

By running this statistical test that allows the R^2 values to be compared, it was found that the linear equation gave a better statistical fit. Most of the previous studies

---

1 A full description of the statistics for all the variables can be found in the Appendix (A-3)
regarding attendance in baseball have also used the linear form in their regressions modeling attendance (Demmert 1973; Noll 1974; Hill, Madura and Zuber 1982; Marcum and Greenstein 1985; Scully 1989; Knowles, Sherony and Haupert 1992; Coffin 1996; Schmidt and Berri 2001). For these reasons, and the fact that the linear form makes the interpretation of the coefficients easier, the chosen functional form for this study was the linear form.

All results that are reported in this paper used the linear form. Ordinary Least Squares (OLS) was used with the EViews statistical software package for all the regression estimates. To begin with, preliminary regressions were run including all of the independent variables that were described previously. Through the course of looking at previous studies, and use the use of contemplative theory, along with the use of statistical testing, the final model was produced. Of all the variables that were described, only four were not used in the final model.

The first variable that was not included was the most controversial omission. The inclusion of a variable measuring stadium capacities has been used in three previous studies regarding attendance in major sports. Borland and Lye (1992) looked at single game attendance in Australian Rules football and found that the size of the stadium played in was significant. Marcum and Greenstein (1985) used stadium capacity as part of their dependent variable. They divided attendance by capacity in order to compare the percentage of the stadium filled at the ballparks of the two teams they studied over the 1982 season: the St. Louis Cardinals and the Texas Rangers. Coffin (1996) found the stadium capacity to be a significant variable in his study of seasonal attendance between 1962-1992. This variable makes sense in terms of compensating for games that sell out.

4 A description of this test can be found in the Appendix (A-4)
These games have an excess demand than what is presented in the attendance figures. However, over the data set that this study uses, only 5% of the games played had sellouts. Also, the other studies that have been done on sports attendance have not included a stadium capacity variable (Demmert 1973; Noll 1974; Hill, Madura and Zuber 1982; Scully 1989; Knowles, Sherony and Haupert 1992; Bruggink and Eaton 1996; Schmidt and Berri 2001). The inclusion of this variable in the model caused many variables to lose their hypothesized signs, even though STDCAP was a statistically significant variable. Due to the fact that this variable has not been dominantly used in previous studies, and that the variable only affects a small percentage of games, the variable STDCAP was not included into the final model.

The variable that measured the number of games out of first place the away team was, ABACK, was also not included in the final model. Although this variable was included in all four of the previous studies on single game attendance in baseball (Hill, Madura and Zuber 1982; Marcum and Greenstein 1985; Knowles, Sherony and Haupert 1992; Bruggink and Eaton 1996), there is a good reason for its omission. These previous studies used the number of games back a team was their only variables regarding team quality. None of the four studies used any type of winning percentage variable for either the home or the away team. When the number of away games behind is added to the model in this study, both of the winning percentage variables become highly insignificant. This is due to the fact that winning percentage and games behind appear to be highly correlated, so dropping one doesn't hurt. By dropping ABACK nothing is lost in the model because the winning percentage variables cover the effects that would have been lost.
The last two variables dropped were \textit{DIVWINDIFP} and \textit{DIVWINDIFN}. These two variables were insignificant at all levels, and were fairly hard to interpret. When added to the model these variables were already difficult to justify putting in by theory. The preliminary regressions realized these misgivings about the variables by finding them to be insignificant. To make sure that the variables were not unduly omitted, the log-likelihood test was run and it found these variables were not statistically different to zero in the model.

\textbf{EMPIRICAL RESULTS}\textsuperscript{5}

\begin{align*}
\text{ATTENDANCE} &= B_1 + B_2 \text{SUN} + B_3 \text{MON} + B_4 \text{TUES} + B_5 \text{WED} + B_6 \text{THUR} + \\
& \quad B_7 \text{FRI} + B_8 \text{CMSAPOP} + B_9 \text{INC} + B_{10} \text{FCI} + B_{11} \text{COMP} + B_{12} \text{DOME} + \\
& \quad B_{13} \text{RETRACT} + B_{14} \text{DAYGAME} + B_{15} \text{DOUBLEH} + B_{16} \text{AVGTEMP} + \\
& \quad B_{17} \text{PRECIP} + B_{18} \text{TWOTHSONE} + B_{19} \text{TWOTHSTWO} + \\
& \quad B_{20} \text{PRVPLYOFF} + B_{21} \text{PRVWS} + B_{22} \text{PRVWSCHAMP} + B_{23} \text{GAMESBACK} + \\
& \quad B_{24} \text{DIV} + B_{25} \text{GAMESLEFT} + B_{26} \text{HALLSTARS} + B_{27} \text{ALLSTARS} + \\
& \quad B_{28} \text{WINDIFFP} + B_{29} \text{WINDIFFN} + B_{30} \text{GLWINDFP} + B_{31} \text{GLWINDFN} + \\
& \quad \text{u}
\end{align*}

Days of the Week

\begin{center}
\textit{SUN, MON, TUES, WED, THUR, FRI, SAT}
\end{center}

\textsuperscript{5} A copy of the final model can be found in the Appendix (A-5)
With the previous studies all finding that the day of the week did affect attendance, the results of this study did not differ. The omitted variable was Saturday; the coefficients of the day variables thus represent the level of attendance relative to Saturday. As expected, Friday and the weekend days strongly outdraw the weekday games. All the days of the week are statistically significant at the .01 level, with Saturday drawing the largest crowd. With Saturday having the greatest effect on attendance of the days, and it being the omitted variable, all the coefficients for the day variables are negative. Ceteris paribus, Sunday games cause attendance to be 3627 less than it would be on Saturday, Fridays are 2542 less, Mondays are 7053 less, Tuesdays-8142 less, Wednesdays-8050, and Thursdays-7645. These results are consistent with what has previously been seen in single game attendance studies. The only study that looks at each day specifically, as opposed to the “weekend” and “midweek” dummy variables, is the Marcum and Greenstein study (1985). Their study looked at the percentage of the stadium filled instead of straight attendance and make comparisons of the coefficients more difficult, but their results seem quite consistent with these numbers. They too found Saturday to have the greatest effect on attendance, followed by Friday and Sunday, then the midweek variables roughly similar. These results coincide with the theory that was used in including the variables.

Market Variables

*CMSPAPOP, INC*

In all previous studies regarding attendance of major sports, both seasonal and single game, the population of the home teams market has been a positive and
statistically significant variable of attendance. That is the same result that this study has found. Ceteris paribus, an extra thousand people in the CMSA population of the home team's metropolitan area will add 0.118 people to attendance. This is more easily grasped when using greater numbers; an increase in 1,000,000 people will cause a games attendance to increase by 118,313 people, ceteris paribus. This variable is statistically significant at the .01 level.

This study lies roughly in the middle of previous studies findings as far as how much the size of the market really adds to attendance. While this variable is positive and significant, it is obviously not a huge determinant of attendance. For example, New York (the largest CMSA in MLB at 21,200 mil.) will only outdraw Milwaukee (the smallest CMSA in MLB at 1,690 mil.) by 2,300 fans, ceteris paribus, even though New York is a vastly larger market.

Per capita income was used in previous studies and was found to be positive and statistically significant towards attendance. The only study that didn't find per capita income to have a positive effect on attendance was Noll (1974); where he hypothesized that baseball could be an inferior good (p. 120). That was not found in this study. A one dollar increase in the average per capita income of a home team's metropolitan area is associated with an increase in 0.243 fans in attendance. This is quite a large amount considering that an increase from the median team in the MLB ($32,515) to the San Francisco Giants (highest in MLB at $57,414) is associated with an increase in 6,050 fans, ceteris paribus. This variable shows that if baseball truly was an inferior good in 1974 when Noll did his study, it isn't anymore.
Home Team Fixed Variables

\[ FCI, \text{COMP} \]

The price of a game is something that can clearly affect the attendance of it. If the game is overpriced, the demand will show it. The previous results of studies are inconclusive with the effect of price on baseball attendance. This study has found that attendance is positively related to attendance and is statistically significant at the .01 level. With a rather large gap between the prices charged by the least expensive team (Montreal at $76.77) and the most expensive team (Boston at $228.73), this is interesting. With \( FCI \) having a coefficient of 111.0535, this represents, ceteris paribus, an increase in almost 17,000 fans per game from Montreal to Boston! This variable is misleading however. The fan cost index is quite highly correlated to the winning percentage, home team All-Stars and per capita income. In baseball, high winning percentage and a high number of All-Stars equates into a team that has a high payroll and a team that is entertaining to watch. The fact that a team has a high payroll forces management to charge higher prices on tickets and complementary goods. \( FCI \) is actually more than 93% correlated with the ticket price of the home team. The fact that a team has a high winning percentage and a lot of All-Stars also makes the demand to attend this teams are greater. Due to greater demand the management can charge higher prices and still maintain attendance. Also, since \( FCI \) is highly correlated with the home team per capita income, management in these higher income cities can charge higher prices and not lose demand, because the inhabitants of those cities make more money, ceteris paribus. These two effects combine to show that a higher quality of team is associated with higher \( FCI \),
so it is not that alarming that the sign on FCI is positive. After all, higher quality goods cost more than lower quality goods across all fields of business.

The presence of another MLB team in one's home CMSA causes attendance to decrease, ceteris paribus. That result is consistent with what previous studies on attendance have found (Demmert 1973; Noll 1974; Hill, Madura and Zuber 1982; Bruggink and Eaton 1996; Garcia and Rodriguez 2001). This study has found that having a competing team in your metro area is associated with a decrease in about 1,400 fans at a game, ceteris paribus, shown by the coefficient of COMP at -1399.496. This result is statistically significant at the .01 level. While this result is expected, it should be noted that the markets where two teams have been placed are very large. The four CMSA's that are affected by this variable are the 1st, 2nd, 3rd, and 5th largest in the nation. The New York area has over 21 million residents, Los Angeles has over 16 million, Chicago over 9 million, and the Bay Area has over 7 million. These markets have two teams because their market(s) can support them. The fourth ranked CMSA is the greater Washington D.C.-Baltimore area at 7.6 million, which has recently been talked about as the vogue place to move the financially struggling Montreal Expos. The fact that these are huge metropolitan areas makes having two teams in them viable; surely most metropolitan areas couldn't support two MLB teams.

Stadium Specific Variables

*DOME, RETRACT*

The atmosphere that a team plays in can have a great effect on attendance. This can clearly be seen in the amazing rise in new facilities, with seventeen new stadiums
opening since the start of the 1989 season and another park scheduled to open next season (2004). The age of a home teams stadium was not included for the reasons stated earlier, however, it is clear that the atmosphere baseball is played in can affect attendance. That was the reasoning for adding DOME, and RETRACT. If a team plays in a domed stadium in can expect to have 9,185 fewer fans than if they played in an outdoor park! It should be noted, however, that only three MLB teams currently play in domed stadiums: the Minnesota Twins, Montreal Expos, and the Tampa Bay Devil Rays; not exactly the who’s-who of the MLB elite. These are small market, low payroll teams that have traditionally been three of the worst drawing teams in baseball (except for the Twins '02 season when they qualified for the ALCS), even when the quality on the field was good. This could be attributed to the stadium that the teams play in, or it could be these cities are poor baseball markets, or that management just doesn’t put a quality team on the field. So while this variable is statistically significant, it should be noted that extenuating circumstances may be at play other than just the home stadium.

If a team plays in a retractable roof stadium they can expect an increase of 1,906 fans over a conventional outdoor park. This variable was statistically significant at the .01 level. The hike in attendance of retractable roof stadiums can most likely be attributed to the fact that these are newer parks that have “fan-friendly” amenities, and they can block out bad weather when necessary. If a fan knows that poor weather is expected for game time, they are likely to skip a game if it is held at a conventional park, but if the game is held in a retractable roof stadium, they know that the roof can be closed and the weather can be kept out. Also, if the weather is nice, the roof can be opened and the day can be enjoyed. Therefore, with a retractable roof stadium the fans get the best of
both worlds, the ability to enjoy the weather when it’s nice, and the ability to close out the weather when it’s bad. This is most likely the reason for the added attendance.

Weather and Game Specific Non-Quality Variables

*DAYGAME, DOUBLEH, AVGTEMP, PRECIP*

Three previous studies have looked at how day games compare in attendance to night games (Hill, Madura and Zuber 1982; Knowles, Sherony and Haupert 1992; Bruggink and Eaton 1996). These three studies were not unanimous in their findings on whether games during the day outdrew games during the night. This study has found that over the period of observation, that a game played during the day can be expected to outdraw a night game by 2,013 fans. *DAYGAME* is statistically significant at the .01 level. Although this variable does not contain the expected sign that was hypothesized, this result was not that surprising. The reason that this variable is positive could be linked to the fact that there was no differentiation between weekend day games and weekday games. It would be expected that weekday day games would have smaller attendance, but that might be offset by an increase in weekend day games. This could possibly be linked with the fact that weekend day games may be more family friendly games to attend. Parents have the weekend free from work, and by going to a day game they don’t have to worry about keeping Junior up past his bedtime.

*DOUBLEH* was not found to be statistically significant at any meaningful level. This somewhat follows with the hypothesis stated earlier in the paper. While the coefficient had a negative coefficient, this is meaningless considering its insignificance as shown by its t-statistic (-0.745). This can most likely be linked to the fact that there are
no longer any double headers scheduled; they are now all the results of games that have been rescheduled due to bad weather or other circumstances necessitating a reschedule. Therefore, fans don’t make a day at the ballpark where they get to see two games with one ticket. Two tickets must now be taken for each game, making each game a separate entity and causing no added incentive to attend either leg of the double header.

The weather on gameday obviously affects a fan’s decision to attend a game. \( \text{AVGTEMP} \) was found to be statistically significant at the .01 level and positive in relation to attendance. For each one degree increase in the average temperature the day a game was played, attendance could be expected to increase by 99 fans, ceteris paribus. That follows with the hypothesized sign that was discussed earlier in the paper. As discussed, at the highest extreme, greater temperature creates a disincentive to go to a game, but baseball is a summer sport made to be played outside. This study shows that fans think the hotter the better.

Rain is never appreciated at an outdoor event, baseball included. This study found that for each inch of rain that falls on the day of the game, 1,539 less fans can be expected, ceteris paribus, than if there were no rain. \( \text{PRECIP} \) was found to be statistically significant at the .01 level. This result is consistent with the prior hypothesis and previous studies: rain causes fans to stay away from the ballpark.

**Yearly Variables**

\( \text{TWO\text{-}THSNDONE, TWO\text{-}THSNDTWO} \)

These two variables clearly have little to report about. The variables were added to control for the fact that this study covers three different seasons, and that different,
unaccounted for, factors may have caused attendance to fall in any of these seasons.

With the omitted variable being games played during the 2000 season, these two
variables give the attendance in their respective seasons compared to the 2000 season.
Since no single game study has covered multiple seasons, there is little to compare the
results with. Ceteris paribus, games played during the 2001 season have 618 less fans
than 2000, and games in 2002 have 2,700 less fans than 2000. Without really studying
the effects behind this I don’t have a definitive answer as to why attendance has declined.
An educated guess is that with the slowdown of the economy the last three years, people
are less likely to spend money going to a baseball game; instead choosing a cheaper
alternative, such as a movie, or catching the game on TV.

Playoff Lagged Variables

$PRVPLYOFF, PRVWS, PRVWSCHAMP$

The success that a team has in years prior their season has been shown to affect
attendance (Demmert 1973; Noll 1974; Hill, Madura and Zuber 1982; Borland and Lye
1992; Bruggink and Eaton 1996). The ways that these studies have chosen to account for
this have differed, but many have used the number of pennants the team has recently
won. Which before each league expanded from two to three divisions was equivalent to a
divisional championship. This study explores how deep a team gets in the playoffs
affects attendance the next year. If a team qualified for the postseason in the previous
season, measured by $PRVPLYOFF$, it can be expected to add 1,489 to each games
attendance the next season over a team that didn’t qualify for the playoffs, ceteris paribus.
This variable was found to be significant at the .01 level. If a team qualified for the
World Series, measured by PRVWS, it was found to be insignificant at the .10 level. Therefore, contrary to previous studies, winning the pennant (each league champion wins "the pennant") has lost its significance in affecting attendance. The diminished importance of winning the pennant can most likely be attributed to the expansion of the playoffs to the three division champions in each league plus a wildcard team. This can be interpreted by the fact that attendance isn’t increased any extra by a playoff team qualifying for the World Series; all the added attendance increase is made by just qualifying for the playoffs. However, if a team wins the World Series, measured by WSCHAMP, a team can expect an increase (on top of the affect qualifying for the playoffs had on attendance) of 1,764 fans per home game the next year, ceteris paribus. These variables have followed with the hypothesis that was made at the beginning of this study, however it is surprising that just being in a World Series game does not greatly affect attendance.

**Single Game Fluctuating Competition Variables**

*GAMESBACK, DIV, GAMESLEFT*

These variables are measurements for the competitiveness of the home team and the meaning that the game has for the home team. The GAMESBACK variable measures the number of games out of first place a team is in its division. Using the number of games a team is out of first place as a variable affecting attendance has a strong history of being used in studies on attendance in sports (Demmert 1973; Noll 1974; Hill, Madura and Zuber 1982; Marcum and Greenstein 1985; Borland 1987; Scully 1989; Knowles, Sherony and Haupert 1992; Borland and Lye 1992; Bruggink and Eaton 1996; Coffin
1996; Garcia and Rodriguez 2001; Schmidt and Berri 2001). These studies have all found that the number of games back of first place has a negative effect on attendance, whether those studies looked at seasons or single games. This study finds the same relationship. With each game that a team is out of first place, it loses 307 fans, ceteris paribus. Therefore a team that was ten games out of first place would have 3,068 less fans at a game than a division leader, ceteris paribus.

With the playoffs being largely decided by how a team finishes in its division, these games become more important. Bruggink and Eaton (1996) have previously looked at how within division games affected attendance on a per game basis. They found that this variable was positive and statistically significant. This study did not come to the same conclusion; the variable DIV was found to be highly insignificant. This could possibly be attributed to the existence of the wildcard position available for qualification into the postseason. By creating an option of making the playoffs without winning your divisional title, it makes the importance of the divisional games to become on par with games against other, non-divisional, opponents. This would be consistent with the fact that this variable was found not to be statistically different than zero.

Due to the length of the baseball season, or any professional spots league for that matter, games have an added meaning when teams are jockeying for playoff berths at the end of the season. The variable measuring how many games remain in the regular season, GAMESLEFT, was found to be negative and statistically significant at a .01 level. The coefficient of this variable can be interpreted as follows: for each game closer to the end of a season a team plays, they can expect an added 35 fans to each game. Or, the last game of the season (the 162nd) will have 5,258 more fans than the 10th game of the
season, ceteris paribus. This variable accounts for the added excitement that is associated with teams making runs for the playoffs, and also accounts for the low drawing early months of April and May.

**Player Quality**

\[ \text{HALLSTARS, AALLSTARS} \]

The existence of stars always has an effect on attendance. The presence of a few stars can have as big an effect on attendance as any other factor. Three of the previous studies on attendance in sports have used some sort of All-Star or star player variables (Noll 1974; Bruggink and Eaton 1996; Garcia and Rodriguez 2001). This study has found the presence of All-Stars to greatly affect attendance. The number of home All-Stars, \( \text{HALLSTARS} \), and the number of away All-Stars, \( \text{AALLSTARS} \), are both positive and statistically significant at the .01 level. An increase of one All-Star on the home team will increase attendance at each game by 1091 fans, ceteris paribus. With the average ticket price in MLB in 2002 being $18.30 (www.teammarketing.com), and multiplying that by the 81 home games that each team plays, one additional All-Star on a team adds over $1.6 million to team revenues in just ticket revenue over a full season! For each away All-Star at a game, 869 additional fans can be expected, ceteris paribus.

**Uncertainty Variables**

\[ \text{WINDIFFP, WINDIFFN} \]

These are the two variables that were critical to this study's intent. Until now all the other variables were used to control for effects on attendance, so that these variables
could specifically be looked at. As far as variables like these being used in previous studies, none could be found—this is a brand new way of looking at competitive balance and estimating attendance. A few previous studies (Scully 1989; Knowles, Sherony and Haupert 1992; Coffin 1996) looked at the winning percentages for the teams, but never a difference between the two. These variables behave the way that was hypothesized and are both statistically significant at the .01 level. The results that have been found here do not support the competitive balance argument.

\( WINDIFFP \), the variable measuring the difference in winning percentage when the home team is better than the away team, is negative as expected. \( WINDIFFP \) has a coefficient of \(-19,935.58 \). This can be interpreted easily: if the home team has a .650 winning percentage (consistent with the elite teams of the MLB) and the visiting team has a .400 winning percentage (consistent with the lower level MLB teams), the home team can expect attendance to be 4,984 less than if the visiting team had the same winning percentage as the home team (.650), ceteris paribus. This value of the coefficient fulfills both effects on attendance, for and against competitive balance. The competitive balance supporters would look at this and say that the reason for the decreased attendance is that the two winning percentages are different, that the uncertainty of outcome has been diminished and therefore would affect attendance poorly. However, for that argument to hold water, \( WINDIFFN \), would also have to have a negative coefficient, but it doesn't.

\( WINDIFFN \), the variable that measures the difference in winning percentage when the home team has a poorer record than the visiting team, doesn't have the negative coefficient that would support the competitive balance theory. \( WINDIFFN \) has a coefficient of 10,138.49. Look at the same example as before, except make the .400 team
the home team and the .650 team the visitors. In this case the home team can expect 2,535 more fans than if the visiting team had the same record (.400), ceteris paribus. This is an extremely significant finding in that it refutes the competitive balance argument. If the competitive balance argument held here, WINDIFFN would have had a negative effect on attendance.

The implications of these results are that attendance is maximized when the opponent is the elite team of the league. Even if the home team has the best record in the league, they maximize attendance by playing the team with the next best record. While this is generally expected, the fact that attendance is maximized for poor teams when the league elite come to town is contrary to what many have hypothesized. This result refutes much of the work done on the uncertainty hypothesis, which argues attendance will be maximized when the home team is favored (Knowles, Sherony and Haupert 1992). In their study they found that attendance was greatest when the home team has a .6 chance of winning. While there is no measure of chance of winning in this study, one can assume that a .400 team will not be favored when playing the best team in the league.

Interaction Terms

GLWINDFP, GLWINDFN

These two terms were the product of the GAMESLEFT variable and the difference in winning percentage variables. As these variables used the winning percentage variables, no other study has used variables like these before. The results were interesting, GLWINDFP was positive and statistically significant at the .01 level, but GLWINDFN was found to be not statistically different than zero at the .10 level. That
result showed that the number of games left affected attendance when the home team was playing a team that was worse than it, but not when it played a team that was better than it. The coefficient on \textit{GLWINDFP} was 168.9707. This can be interpreted that at the beginning of the season fans will be more apt to go to a game involving an inferior team than at the end of the season. For example: in the 10\textsuperscript{th} game of the season (152\textsuperscript{nd} remaining) the .650 team can expect 6,421 more fans, ceteris paribus. In the 152\textsuperscript{nd} game (with 10 games remaining), the same scenario of teams will only add 422 fans. This can be looked at as fans becoming more selective in the opponent as the season continues. A team that is bad at the beginning of the season may just be in a slump, having the potential to turn around a bad season. However, a team that is bad at the end of the season is most likely just plain bad. The fact that this relationship doesn't hold for when the visiting team is better than the home team just shows that the time of the season doesn't affect attendance when a superior team comes to town.

\textbf{CONCLUSION}

\textbf{Implications}

This study has shown the key variables that determine the demand for single games in Major League Baseball. More important than that, however, is that this study has shown that what was previously thought about competitive balance to be false. At the very least, this study raises some serious doubts to the assumption that the greater the competitive balance of a game, the greater attendance will be. This study has shown that
attendance for a single game is maximized when the greatest quality, visiting opponent is played. It has been shown to be true even when the inferior, home team has little chance of winning. This could be related to how our culture often roots for the underdog; that fans show up to games on the chance that a great upset could be seen. It could also be attributed to the fact that fans enjoy seeing the very elite that a sport has to offer. Either way, attendance is maximized when the highest quality visiting team is played.

The fact that competitive balance does not maximize attendance at games, per se, does not mean that it doesn’t maximize league wide attendance. The fact that the negative effect on attendance that WINDIFFP has is greater than the positive effect on attendance that WINDIFFN causes competitive balance to maximize league wide attendance. WINDIFFP possesses a coefficient of almost —20,000, while WINDIFFN only possesses a coefficient of about 10,000. The implications of this is that while a .400 team hosting a .650 team at home can expect attendance to increase by 2,500, but the .650 team hosting that same .400 team can expect attendance to decrease by 5,000. The net of these two effects is to decrease league attendance by 2,500. Since this is inevitable, with the inferior teams visiting the superior teams as much as the superior teams visit the inferior teams, the league wide effect on attendance will be negative. With perfect competitive balance, where all teams have a .500 winning percentage, no negative effects would be able to outweigh positive effects, thereby maximizing league attendance. Therefore, it is in the leagues best interest to maximize competitive balance, but for each team, it is their best interest to bring the highest quality opponents to the ballpark as often as possible.
Suggestions for Future Research on the Subject

This study has rather definitively answered the question of how competitive balance affects single game attendance in Major League Baseball. However, that doesn’t mean this question can’t still be explored. For future studies, a variable measuring whether promotions were given away on gameday could be used. The rise of bobbleheads in recent years, and the time tested promotions of bats and helmets have always affected attendance positively. A variable was not included in this study due to the difficulty of obtaining this information and the time constraints put on the author (as this paper is an honors thesis, only roughly 9 calendar months were available to finish this study, on top of other scholastic activities). A number of previous studies have looked at the presence of major and minor promotions on attendance, and have found the effects to be positive and statistically significant (Hill, Madura and Zuber 1982; Marcum and Greenstein 1985; Bruggink and Eaton 1996).

Also, the effect of competitive balance on single game attendance could be looked at for other professional sports. Football may not be a great data set for this type of study due to the fact that there are few games during the season and most of them sell out; and therefore leaving a large amount of excess demand that could not be measured. However, a study of this nature could very well work for the NHL or the NBA, where there are a greater number of games and fewer sellouts than the NFL.
Closing Remarks

In conclusion I would like to thank Randy Nelson for his help as an advisor on things other than just this study. I would also like to thank Jim Meehan and Cliff Reid for their help in turning over variables that may have been overlooked.

I could never have attended Colby College or accomplished anything without the support and love of my family, for whom I dedicate this study. In retrospect, the past four years have been an unbelievable experience that has given me the skills to be successful in the next phase of my life. I thank the great friends that I have made and the Professors that so clearly care for their students for this. It is easy as a second semester “jaded” senior to lament about how I spent $100,000 more on this education than I would have at my state college, but I know the relationships that I have been able to have with faculty, and not just those in my major, has made this well worth the price-tag. I am sad to leave the confines of Mayflower Hill, but am excited about the life that is ahead of me.

-Tom Richardson ’03
References


# Appendix A-1: Variable Descriptions

<table>
<thead>
<tr>
<th>Variable</th>
<th>Description</th>
<th>Exp. Sign</th>
</tr>
</thead>
<tbody>
<tr>
<td>ATTENDANCE</td>
<td>The number of tickets sold at each game</td>
<td></td>
</tr>
<tr>
<td>SUN</td>
<td>Dummy variable equals one if game is played on a Sunday game</td>
<td></td>
</tr>
<tr>
<td>MON</td>
<td>Dummy variable equals one if game is played on a Monday</td>
<td></td>
</tr>
<tr>
<td>TUES</td>
<td>Dummy variable equals one if game is played on a Tuesday</td>
<td></td>
</tr>
<tr>
<td>WED</td>
<td>Dummy variable equals one if game is played on a Wednesday</td>
<td></td>
</tr>
<tr>
<td>THUR</td>
<td>Dummy variable equals one if game is played on a Thursday</td>
<td></td>
</tr>
<tr>
<td>FRI</td>
<td>Dummy variable equals one if game is played on a Friday</td>
<td></td>
</tr>
<tr>
<td>SAT</td>
<td>Dummy variable equals one if game is played on a Saturday</td>
<td></td>
</tr>
<tr>
<td>CMSAPOP</td>
<td>2000 CMSA population (in thousands) of the home team</td>
<td>+</td>
</tr>
<tr>
<td>INC</td>
<td>2000 per capita income of the home team's metropolitan area</td>
<td>+</td>
</tr>
<tr>
<td>FCI</td>
<td>Fan Cost Index of each team for the applicable season</td>
<td></td>
</tr>
<tr>
<td>COMP</td>
<td>Dummy variable equal to one if two teams play in the home teams</td>
<td></td>
</tr>
<tr>
<td>DOME</td>
<td>Dummy variable equal to one if the home team plays in a domed stadium</td>
<td></td>
</tr>
<tr>
<td>RETRACT</td>
<td>Dummy variable equal to one if the home team plays in a stadium with a retractable roof</td>
<td></td>
</tr>
<tr>
<td>DAYGAME</td>
<td>Dummy variable equal to one if the game is played before 5:00 p.m.</td>
<td></td>
</tr>
<tr>
<td>DOUBLEH</td>
<td>Dummy variable equal to one if the game is part of a doubleheader</td>
<td>undet.</td>
</tr>
<tr>
<td>AVGTEMP</td>
<td>Average of the high and low temperature on the day of the game</td>
<td>+</td>
</tr>
<tr>
<td>PRECIP</td>
<td>Amount of rain that falls on the day of the game (in inches)</td>
<td></td>
</tr>
<tr>
<td>TWOOTHSDONE</td>
<td>Dummy variable equal to one if the game was played during the 2001 season</td>
<td></td>
</tr>
<tr>
<td>TWOOTHSDTWO</td>
<td>Dummy variable equal to one if the game was played during the 2002 season</td>
<td>undet.</td>
</tr>
<tr>
<td>PRVPLYOFF</td>
<td>Dummy variable equal to one if the home team qualified for the playoffs</td>
<td>+</td>
</tr>
<tr>
<td>PRVWS</td>
<td>Dummy variable equal to one if the home team qualified for the World Series</td>
<td>+</td>
</tr>
<tr>
<td>PRVWSCHAMP</td>
<td>Dummy variable equal to one if the home team won the World Series</td>
<td>+</td>
</tr>
<tr>
<td>GAMESBACK</td>
<td>Number of games that the home team is behind the first place team in their division</td>
<td></td>
</tr>
<tr>
<td>DIV</td>
<td>Dummy variable equal to one if the home and visiting teams are from the same division</td>
<td></td>
</tr>
<tr>
<td>GAMESLEFT</td>
<td>Number of games left in the season, (162-(wins+losses))</td>
<td></td>
</tr>
<tr>
<td>HALLSTARS</td>
<td>Number of All-Stars on the home team's roster</td>
<td>+</td>
</tr>
<tr>
<td>AALLSTARS</td>
<td>Number of All-Stars on the visiting team's roster</td>
<td>+</td>
</tr>
<tr>
<td>WINDIFFP</td>
<td>Difference of two teams winning percentage when home team is better</td>
<td></td>
</tr>
<tr>
<td>WINDIFFN</td>
<td>Difference of two teams winning percentage when visiting team is better</td>
<td></td>
</tr>
<tr>
<td>GLWINDFP</td>
<td>Interaction term equal to (GAMESLEFT*WINDIFFP)</td>
<td>+</td>
</tr>
<tr>
<td>GLWINDFN</td>
<td>Interaction term equal to (GAMESLEFT*WINDIFFN)</td>
<td>-</td>
</tr>
</tbody>
</table>
Appendix A-2: \textit{WINDIFF} variables helper chart

\[ \text{WINDIFF} = \text{Home Team Winning \%} - \text{Visiting Team Winning \%} \]

\textit{WINDIFF} is not monotonic.
# Appendix A-3: Descriptive Statistics of the Variables

<table>
<thead>
<tr>
<th>Variable</th>
<th>Mean</th>
<th>Median</th>
<th>Maximum</th>
<th>Minimum</th>
<th>Std. Dev.</th>
</tr>
</thead>
<tbody>
<tr>
<td>ATTENDANCE</td>
<td>29451.55</td>
<td>30334</td>
<td>61277</td>
<td>2134</td>
<td>11792.49</td>
</tr>
<tr>
<td>SUN</td>
<td>0.159663</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0.366318</td>
</tr>
<tr>
<td>MON</td>
<td>0.091098</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0.287768</td>
</tr>
<tr>
<td>TUES</td>
<td>0.150539</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0.357624</td>
</tr>
<tr>
<td>WED</td>
<td>0.154271</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0.361234</td>
</tr>
<tr>
<td>THUR</td>
<td>0.123307</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0.328812</td>
</tr>
<tr>
<td>FRI</td>
<td>0.159248</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0.365932</td>
</tr>
<tr>
<td>SAT</td>
<td>0.161874</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0.368361</td>
</tr>
<tr>
<td>INC</td>
<td>33417.13</td>
<td>32515</td>
<td>57414</td>
<td>21160</td>
<td>6364.653</td>
</tr>
<tr>
<td>FCI</td>
<td>139.4994</td>
<td>138.33</td>
<td>228.73</td>
<td>76.77</td>
<td>27.28926</td>
</tr>
<tr>
<td>CMSAPPOP</td>
<td>6308.374</td>
<td>4112</td>
<td>21200</td>
<td>1690</td>
<td>5348.059</td>
</tr>
<tr>
<td>COMP</td>
<td>0.266243</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0.442023</td>
</tr>
<tr>
<td>DAYGAME</td>
<td>0.291955</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0.454693</td>
</tr>
<tr>
<td>DOUBLEH</td>
<td>0.014515</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0.119608</td>
</tr>
<tr>
<td>AVGTMP</td>
<td>69.01009</td>
<td>70.5</td>
<td>101</td>
<td>30.5</td>
<td>10.81874</td>
</tr>
<tr>
<td>PRECIP</td>
<td>0.076315</td>
<td>0</td>
<td>5.4</td>
<td>0</td>
<td>0.272176</td>
</tr>
<tr>
<td>DOME</td>
<td>0.100221</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0.300315</td>
</tr>
<tr>
<td>RETRACT</td>
<td>0.167127</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0.373115</td>
</tr>
<tr>
<td>WINDIFFP</td>
<td>0.063328</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0.114555</td>
</tr>
<tr>
<td>WINDIFFN</td>
<td>0.060366</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0.108803</td>
</tr>
<tr>
<td>GAMESBACK</td>
<td>8.143558</td>
<td>5.5</td>
<td>45</td>
<td>0</td>
<td>8.483641</td>
</tr>
<tr>
<td>TWOTHSND</td>
<td>0.333564</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0.471519</td>
</tr>
<tr>
<td>TWOTHSDONE</td>
<td>0.333011</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0.471323</td>
</tr>
<tr>
<td>TWOTHSDTWO</td>
<td>0.333425</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0.47147</td>
</tr>
<tr>
<td>DIV</td>
<td>0.416091</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0.492943</td>
</tr>
<tr>
<td>GAMESLEFT</td>
<td>80.77067</td>
<td>81</td>
<td>161</td>
<td>0</td>
<td>46.64569</td>
</tr>
<tr>
<td>GLWINDFP</td>
<td>6.248225</td>
<td>0</td>
<td>161</td>
<td>0</td>
<td>16.30151</td>
</tr>
<tr>
<td>GLWINDFN</td>
<td>5.83702</td>
<td>0</td>
<td>161</td>
<td>0</td>
<td>15.19859</td>
</tr>
<tr>
<td>PRVPLYOFF</td>
<td>0.266934</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0.442388</td>
</tr>
<tr>
<td>PRVWS</td>
<td>0.066768</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0.249637</td>
</tr>
<tr>
<td>PRVWSCHAMP</td>
<td>0.033315</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0.17947</td>
</tr>
<tr>
<td>HALLSTARS</td>
<td>2.139895</td>
<td>2</td>
<td>8</td>
<td>0</td>
<td>1.525235</td>
</tr>
<tr>
<td>AALSTARS</td>
<td>2.132568</td>
<td>2</td>
<td>8</td>
<td>0</td>
<td>1.525572</td>
</tr>
<tr>
<td>DIVWINDFP</td>
<td>0.02821</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0.086233</td>
</tr>
<tr>
<td>DIVWINDFN</td>
<td>0.027507</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0.085152</td>
</tr>
<tr>
<td>ABACK</td>
<td>8.203069</td>
<td>5.5</td>
<td>48</td>
<td>0</td>
<td>8.502719</td>
</tr>
<tr>
<td>STDCAP</td>
<td>47194.95</td>
<td>47000</td>
<td>62409</td>
<td>33871</td>
<td>6207.573</td>
</tr>
</tbody>
</table>
Appendix A-4: Statistical test for functional form

1) Run linear equation- get $R^2$
2) Run ln(y) equation
3) obtain fitted values of log $\hat{y}$
4) Compute $\exp(\log \hat{y}) = \hat{m}$
5) Regress y on $\hat{m}$, with no intercept term, call this new coefficient $\hat{\alpha}_0$
6) Compute $\hat{y} = \hat{\alpha}_0 \times \hat{m}_0$
7) Compute correlation between $\hat{y}$ and y
8) square the correlation- this number (the altered $R^2$ of the log-linear equation) can then be compared with the $R^2$ of the linear equation

Using this test it was found that the comparable $R^2$ value of the log-linear equation was at roughly .47, whereas the $R^2$ of the linear function is .50. Therefore the linear function is used.
Appendix A-5: Final Model

Dependent Variable: ATTENDANCE
Method: Least Squares
Sample: 1 7279
Included observations: 7234
Excluded observations: 45

<table>
<thead>
<tr>
<th>Variable</th>
<th>Coefficient</th>
<th>Std. Error</th>
<th>t-Statistic</th>
<th>Prob.</th>
</tr>
</thead>
<tbody>
<tr>
<td>C</td>
<td>5513.851</td>
<td>1322.575</td>
<td>4.169028</td>
<td>0.0000</td>
</tr>
<tr>
<td>SUN</td>
<td>-3627.273</td>
<td>368.4204</td>
<td>-9.845471</td>
<td>0.0000</td>
</tr>
<tr>
<td>MON</td>
<td>-7053.781</td>
<td>411.8619</td>
<td>-17.12657</td>
<td>0.0000</td>
</tr>
<tr>
<td>TUES</td>
<td>-8142.055</td>
<td>363.1855</td>
<td>-22.41844</td>
<td>0.0000</td>
</tr>
<tr>
<td>WED</td>
<td>-8050.238</td>
<td>355.8050</td>
<td>-22.62542</td>
<td>0.0000</td>
</tr>
<tr>
<td>THUR</td>
<td>-7645.043</td>
<td>370.7935</td>
<td>-20.61806</td>
<td>0.0000</td>
</tr>
<tr>
<td>FRI</td>
<td>-2542.491</td>
<td>358.2385</td>
<td>-7.097202</td>
<td>0.0000</td>
</tr>
<tr>
<td>CMSAPPOP</td>
<td>0.118313</td>
<td>0.041414</td>
<td>2.856846</td>
<td>0.0043</td>
</tr>
<tr>
<td>INC</td>
<td>0.243129</td>
<td>0.024335</td>
<td>9.990755</td>
<td>0.0000</td>
</tr>
<tr>
<td>FCI</td>
<td>111.0535</td>
<td>5.643719</td>
<td>19.67735</td>
<td>0.0000</td>
</tr>
<tr>
<td>COMP</td>
<td>-1399.496</td>
<td>504.7513</td>
<td>-2.772644</td>
<td>0.0056</td>
</tr>
<tr>
<td>DOME</td>
<td>-9185.345</td>
<td>369.3980</td>
<td>-24.86571</td>
<td>0.0000</td>
</tr>
<tr>
<td>RETRACT</td>
<td>1905.639</td>
<td>290.7071</td>
<td>6.555187</td>
<td>0.0000</td>
</tr>
<tr>
<td>DAYGAME</td>
<td>2012.588</td>
<td>277.3375</td>
<td>7.256816</td>
<td>0.0000</td>
</tr>
<tr>
<td>DOUBLEH</td>
<td>-615.8195</td>
<td>826.7231</td>
<td>-0.744892</td>
<td>0.4564</td>
</tr>
<tr>
<td>AVGTEMP</td>
<td>99.14076</td>
<td>10.61007</td>
<td>9.344025</td>
<td>0.0000</td>
</tr>
<tr>
<td>PRECIP</td>
<td>-1539.015</td>
<td>363.2586</td>
<td>-4.236692</td>
<td>0.0000</td>
</tr>
<tr>
<td>TWOTHSNDONE</td>
<td>-618.4070</td>
<td>249.3521</td>
<td>-2.480055</td>
<td>0.0132</td>
</tr>
<tr>
<td>TWOTHSNDTWO</td>
<td>-2699.924</td>
<td>256.0430</td>
<td>-10.54481</td>
<td>0.0000</td>
</tr>
<tr>
<td>PRVPLYOFF</td>
<td>1489.463</td>
<td>275.5687</td>
<td>5.405052</td>
<td>0.0000</td>
</tr>
<tr>
<td>PRVWS</td>
<td>-1017.238</td>
<td>640.0022</td>
<td>-1.589429</td>
<td>0.1120</td>
</tr>
<tr>
<td>PRVWSCAMP</td>
<td>1764.149</td>
<td>768.0494</td>
<td>2.296922</td>
<td>0.0217</td>
</tr>
<tr>
<td>GAMESBACK</td>
<td>-306.8273</td>
<td>18.97597</td>
<td>-16.16925</td>
<td>0.0000</td>
</tr>
<tr>
<td>DIV</td>
<td>-146.2369</td>
<td>203.5194</td>
<td>-0.718540</td>
<td>0.4724</td>
</tr>
<tr>
<td>GAMESLEFT</td>
<td>-34.59361</td>
<td>3.395239</td>
<td>-10.18886</td>
<td>0.0000</td>
</tr>
<tr>
<td>HALLSTARS</td>
<td>1090.755</td>
<td>78.36180</td>
<td>13.91948</td>
<td>0.0000</td>
</tr>
<tr>
<td>AALLSTARS</td>
<td>869.1017</td>
<td>70.40069</td>
<td>12.34507</td>
<td>0.0000</td>
</tr>
<tr>
<td>WINDIFFP</td>
<td>-19935.58</td>
<td>3244.814</td>
<td>-6.143829</td>
<td>0.0000</td>
</tr>
<tr>
<td>WINDIFFN</td>
<td>10138.49</td>
<td>3637.809</td>
<td>2.786978</td>
<td>0.0053</td>
</tr>
<tr>
<td>GLWINDFP</td>
<td>168.9707</td>
<td>22.76092</td>
<td>7.423720</td>
<td>0.0000</td>
</tr>
<tr>
<td>GLWINDFN</td>
<td>-40.76937</td>
<td>25.28916</td>
<td>-1.612128</td>
<td>0.1070</td>
</tr>
</tbody>
</table>

R-squared : 0.504088
Adjusted R-squared : 0.502023
S.E. of regression : 8321.666
Sum squared resid : 4.99E+11
Log likelihood : -75547.62
Durbin-Watson stat : 0.790952

Mean dependent var : 29451.55
S.D. dependent var : 11792.49
Akaike info criterion : 20.89539
Schwarz criterion : 20.92490
F-statistic : 244.0589
Prob(F-statistic) : 0.000000