

Keep Your Eye on the Game: The Impact of Distraction and Scoreboard Watching during  
Major League Baseball Playoff Races

Mike Ferguson, BBA

Submitted in partial fulfillment of the requirements for the degree of  
Master of Arts in Applied Health Sciences  
(Sport Management)

Faculty of Applied Health Sciences

Brock University

St. Catharines, ON

© September 2017

## Abstract

This study examines 11 years and 26,370 of Major League Baseball's (i.e., MLB) game outcomes to test whether distraction, through scoreboard watching, causes teams to choke under pressure. Results indicate that scoreboard watching significantly impacts the probability of winning a game, especially in playoff races. Specifically, teams in a playoff race had a 0.158 lower probability of winning games when the division leader won its game the previous day. Consistent with distraction theory, the analysis also shows that the distraction effects are 0.224 greater on home teams. There is evidence of increased distraction as criticality of games increase. When there are fewer than 10 games remaining in a playoff race, the impact of a division leader win reduces a team's win probability by 0.243. Changes to league structure reduced win probability by 0.039 for seasons starting in 2012. This involved the addition of a Wild Card team to each league and an increase to the value of winning a division.

This study helps fill a gap in the literature in relation to research on external factors and their impact on game outcomes. If a team can account for factors related to winning a game then it could be possible to gain a competitive advantage over the opposition. The findings also have practical applications. MLB teams can take initiatives to eliminate distraction and keep players' attention on the task at hand surrounding critical games.

**Keywords:** Distraction, Scoreboard Watching, Major League Baseball

## **Acknowledgements**

I would like to dedicate my research to my parents, John and Cheryl Ferguson, my grandparents, Joyce and the late Eric Brown, my sister, Pam, and my niece, Hailie. Each of you have understood how much of a passion for sport I have had throughout my life and always been fully supportive. You have been with me through the good times and the bad times and have made me a better person by being such a large part of my life.

My parents knew that I was at a point in my life that I was not fully satisfied with my employment and instead of telling me to stick with it; they said that if I did not go for my dreams then they might not happen. They always ensured that I got to all of my games, whether it was baseball, curling, golf, bowling, or any other sport I played. They also knew of my passion for professional sport and ensured that I got to see many games.

To my grandparents, you both helped raise me as my parents were working full time jobs. You taught me very important life lessons, while still always ensuring that there was plenty of fun. My grandfather would always come to lots of my games and we always had great conversations on many happenings in sport. He has been at rest since February of 2013, but he is definitely still with me. My grandmother still makes sure to talk to me as much as possible and we always have a good laugh and keep up to date.

To my sister, we have not always gotten along perfectly through time, but we have become closer as the years have gone by. I am proud of the person you have become and we will always be there for each other. To my niece Hailie, we are so far away but the simple “Hi’s” from you mean so much and each time seeing you is a special memory.

There are also several other people that I would like to thank for making this possible. I would like to thank my advisor, Dr. Kevin Mongeon, for all of his assistance

throughout this thesis project. You have made me a better researcher and have helped improve my thinking process as a whole. You have always been around to assist with any questions or issues I run into, but also ensure that I do the work on my own to learn from my mistakes. You made sure to make me feel better when I was overstressed.

I would also like to thank the rest of the Sport Management Department, especially my committee of Dr. Chris Chard and Dr. Kirsty Spence. You are always accessible and willing to help me out with any of my questions and even just being around to discuss sport. Your vast areas of expertise have helped grow my overall knowledge of sport. The constant availability, especially of fellow Nova Scotian Charlene, has made my life easier whenever I run into issues.

Thank you to each of the Sport Management graduate students. I knew nobody when I first moved to St. Catharines, and you were always around to chat and just get accustomed to my new life. Ryan, Dan, Nicole, Kari, and Hillary: Each of you gave helpful advice for succeeding in the program and incorporated lots of cool stuff to make grad school traditions. Then, of course, is “The Elitists”. We came in from many different backgrounds, but that first class of 5P08 we said we would have each other’s backs no matter what and I think that has definitely been the case and will continue to be so.

Weller, Michelle, Michaela, and Katelyn: We are all very different people, but I think that just makes us a super team when all our strengths are put together. The future is very bright for each of you. Thank you to Stephanos, Mitch C, Cole, Mitch M, Eddie, Lindsay, Evan, Sean, Tyler, and Chris for all of the entertaining times. Keep up the good work.

I would lastly like to thank all of my friends from home, who are now living in a wide variety of places. The visits, jokes, and chats have meant a lot to me.

Table of Contents

<b>CHAPTER I: INTRODUCTION .....</b>	<b>1</b>
<b>Statement of the Problem.....</b>	<b>1</b>
<b>Purpose.....</b>	<b>3</b>
<b>Research Questions .....</b>	<b>4</b>
<b>CHAPTER II: THEORETICAL BACKGROUND .....</b>	<b>5</b>
<b>CHAPTER III: RELATED LITERATURE .....</b>	<b>8</b>
<b>Choking Under Pressure .....</b>	<b>8</b>
<b>Basketball.....</b>	<b>9</b>
<b>Golf .....</b>	<b>10</b>
<b>Baseball .....</b>	<b>11</b>
<b>Soccer.....</b>	<b>12</b>
<b>Tennis .....</b>	<b>13</b>
<b>Summary of Literature.....</b>	<b>13</b>
<b>CHAPTER IV: EMPIRICAL APPROACH .....</b>	<b>14</b>
<b>Context: Major League Baseball.....</b>	<b>14</b>
<b>Hypotheses .....</b>	<b>15</b>
<b>Data Collection .....</b>	<b>16</b>
<b>Estimation Procedure .....</b>	<b>17</b>
<b>CHAPTER V: RESULTS.....</b>	<b>19</b>
<b>Summary Analysis .....</b>	<b>19</b>
<b>September and October .....</b>	<b>21</b>
<b>August.....</b>	<b>22</b>

<b>July.....</b>	<b>22</b>
<b>June.....</b>	<b>22</b>
<b>Estimation Results .....</b>	<b>23</b>
<b>Entire Season .....</b>	<b>23</b>
<b>Playoff Races.....</b>	<b>29</b>
<b>Team Fixed Effects.....</b>	<b>33</b>
<b>Game Location.....</b>	<b>36</b>
<b>Wild Card.....</b>	<b>37</b>
<b>Games Remaining .....</b>	<b>38</b>
<b>Roster Age.....</b>	<b>39</b>
<b>Team Payroll.....</b>	<b>42</b>
<b>Managerial Experience .....</b>	<b>44</b>
<b>Momentum.....</b>	<b>46</b>
<b>CHAPTER VI: DICSUSSION.....</b>	<b>51</b>
<b>General Findings .....</b>	<b>51</b>
<b>Additional Findings.....</b>	<b>56</b>
<b>Hypothesis test summary .....</b>	<b>58</b>
<b>Limitations .....</b>	<b>58</b>
<b>Areas for Future Research .....</b>	<b>60</b>
<b>CHAPTER VII: CONCLUSION.....</b>	<b>64</b>
<b>REFERENCES.....</b>	<b>67</b>
<b>APPENDIX A: PYTHON CODE FOR ANALYSIS .....</b>	<b>71</b>

## CHAPTER I

### INTRODUCTION

“In investing, just as in baseball, to put runs on the scoreboard one must watch the playing field, not the scoreboard”.

--Warren Buffett<sup>1</sup>

#### **Statement of the Problem**

Traditional sports economic models suggest that game outcomes are determined by relative team qualities and effort, and that players exert effort in response to incentives. Their performance improves with increases to return to effort. An emerging area of research examines the impact of behavioural factors on sports outcomes (Berger & Pope, 2011; Pope & Schweitzer, 2011; Dohmen, 2008; Cao, Price & Stone, 2011), suggesting that the predictions made in classic sports economic models differ systematically from actual observations.

Sports observers have conveyed a popular belief that players' performance often deteriorates when the incentive to perform is greatest. It is typically stated that they “choke under pressure”. Baumeister (1984, p. 391) defines choking under pressure as “performance decrements under circumstances that increase the importance of good or improved performance.” Clark, Tofler and Lardon (2005) and Beilock and Gray (2007) extend the concept by explaining that for a performance to be considered as a choke, a rational athlete must be motivated and capable, and must exhibit the inability to execute due to intervening psychological factors.

---

<sup>1</sup> Warren Buffett is an American business magnate and investor (Szramiak, 2016)

This study examines Major League Baseball's (MLB) high-stakes playoff races to test whether distraction, through scoreboard watching, causes teams to choke under pressure. The term *scoreboard watching* is used to denote a baseball team's awareness of the game outcomes of other teams deemed important to their likelihood of qualifying for the playoffs. As noted by a former baseball player with 17 years of major league experience, the players are actively aware of the scores through the league during playoff races:

“I don't know how confident I was, coming into the last month, thinking we would be able to do it. But every night we found a way to eke out a win. It was fun. We did a lot of scoreboard watching, every night as you take your position in the field, to see what the other scores were.”<sup>2</sup>

Major League Baseball is a well-suited context for examining scoreboard watching and to test for distraction effects. Each season consists of many games that involve the random assignment of competing team qualities. Moreover, previous research has shown that MLB's labour market is efficient and that team payrolls are an excellent proxy for quality. To eliminate possible reverse-causality issues, the research design examines conditional team performance in terms of the division leader's previous day game outcome. Finally, the statistical model isolates the scoreboard watching effect by incorporating a playoff qualification rule change that exogenously alters the value of winning a division.

---

<sup>2</sup> Jeff Conine was a member of the 2003 World Series champion Florida Marlins (Schlossberg, Baxter & Conine, 2004).



## **Purpose**

The purpose of the study was to determine whether the phenomenon of scoreboard watching causes a team to underperform. To this end, the study explored MLB team performance from teams in seasons from 2004 to 2014, and how their performance was impacted by their division leader's previous day game outcome, if at all. Specifically, the number of games remaining in the season and the team's rank prior to the start of a game were examined to determine whether game importance, in terms of the likelihood of qualifying for the playoffs, impacts game outcomes. Moreover, the analysis examined the scoreboard effect under various scenarios, such as the month of the year the game was played, the game location, the playoff structure, and the games remaining.

The study contributes to the growing literature related to the impact of external factors on game outcomes. While previous research examines the impact of momentum and home field advantage, no study thus far examines scoreboard watching.

In terms of managerial implications, teams can attempt to acquire players who have had consistent success in playoff races. Additions of players that have exhibited leadership qualities will help to maintain focus on the task at hand of in September playoff races. Evaluating the ways that the managers and coaching staff are handling situations in game and pre/postgame could be done to teach players to avoid focus on uncontrollable factors such as scoreboard watching. Presentation of study results proving the scoreboard watching phenomenon exists could help motivate players to exert their efforts fully on their performance instead of partial focus on external teams competing.

Each consideration to maximize win probability can help a team achieve additional wins and enhance the likelihood of increased revenue through postseason qualification.

### **Research Questions**

This study investigated the following research questions:

1. Does scoreboard watching negatively impact team performance?
2. Does scoreboard watching cause teams in the playoff race to be more likely to lose a game than teams out of the race?
3. Does scoreboard watching have more of an effect on teams when they play at home or when they play on the road?
4. Did the 2012 change to the Wild Card structure increase the impact of scoreboard watching?
5. Do team roster age levels, team payroll, managerial experience, and momentum contribute to the scoreboard watching effect?

These research questions were developed by considering factors that make Major League Baseball Playoff Races a proper context to evaluate the scoreboard watching phenomenon. This led to finding aspects that differentiated MLB from other leagues such as the change in league structure, the uniqueness of a daily schedule, as well as league rules in regards to league composition. Having a method to test for the scoreboard watching effect was also a requirement. Several statistical tests were created to ensure robust results and avoid the possibility of having a statistical artifact.

## **CHAPTER II**

### **THEORETICAL BACKGROUND**

Choking under pressure can be described as suboptimal performance in relation to what would be expected in a regular situation. This could be brought on by stressful situations that cause psychological concerns in an individual (Baumeister, 1984; Baumeister & Showers, 1986). An example of causation for choking under pressure would be distraction (Wine, 1971), which is the result of attentional disturbances that lead to anxiety.

There are two primary psychological theories that describe reduced performance due to pressure. First, distraction theory (Wine, 1971; Sanders, 1981) suggests that pressure can draw attention to external stimuli (e.g., worry), which causes a reduction in attention on the objective; while second, explicit monitoring theory (Beilock & Carr, 2001) suggests that pressure can increase attention on the objective task to counter-productive levels. Distraction and explicit monitoring theories have traditionally been considered competing perspectives. Specifically, while distraction theory suggests that performance decrements are a result of decreased attention, explicit monitoring suggests the opposite, that they are a result of increased attention.

Sanders and Walia (2012) show that the two theories are not necessarily mutually exclusive. In some cases, decreases in performance can be attributed to the indirect impact of distraction and explicit monitoring on effort rather than entirely on their direct impacts on production. This finding challenges the traditional perspective that the negative effects of pressure are unrelated to effort (Baumeister & Showers, 1986). The literature collectively offers three main conclusions on the causes of choking under

pressure, including: 1) that it is caused by shifts in focus to external stimuli; 2) that it results from an inefficient use of effort; and/or 3) that it is caused by players' over-attention to the consequences of failure.

From a statistical perspective, it can be difficult to isolate the pressure effects to test whether pressure negatively affects outcomes. The pressure of the situation can potentially make the task more difficult to accomplish while decrements in performance can be incorrectly attributed to choking. For example, Gómez, Lorenzo, Jiménez, Navarro and Sampaio (2015) found that winning basketball teams had better performance statistics than losing basketball teams during pressure situations (i.e., close games during the last five minutes of play). It is difficult to attribute these findings as evidence of choking under pressure, however, because winning teams, on average, have better performance statistics than losing teams.

This study examines the impact of choking under pressure through the context of scoreboard watching in MLB. Scoreboard watching is a good context in which to examine the impact of choking under pressure because previously revealed game outcomes are exogenously determined and they do not impact a team's current game strategy. However, scoreboard watching is an unobservable latent variable that cannot be directly measured. In this study, each team's division leader game outcome occurring the previous day is used as a proxy for scoreboard watching. The division leader's game outcomes impact the likelihood that a team will qualify for the postseason and are exogenously determined. They also present a potential source of distraction that can impact another competing team's likelihood to qualify for the playoffs (i.e., the number

of games a team is behind the division leader and the number of games remaining in the season).

## **CHAPTER III**

### **RELATED LITERATURE**

This thesis contributes to the body of research that examines the influence of external factors on team performance. More specifically, the thesis contributes to studies focusing on the phenomenon of choking under pressure. In this chapter, the previous literature that relates to choking under pressure is summarized, and then categorized by sports. For each sport, the studies based on experimental research designs are discussed first, followed by studies based on real world outcomes.

#### **Choking Under Pressure**

The study of choking under pressure attempts to explain the cause of decreased performance in pressured situations. Beyond the world of sports, studies have found evidence of choking under pressure during academic exams (Beilock, Kulp, Holt, & Carr, 2004), and while driving (Fairclough, Tattersall, & Houston, 2006). Teams and/or players in the sport of basketball (Cao, Price, & Stone, 2011; Gómez, Lorenzo, Jiménez, Navarro, & Sampaio, 2015; Mesagno, Marchant, & Morris, 2009; Toma, 2015; Worthy, Markman, & Maddox, 2009), in golf (Beilock, 2007; Hickman & Metz, 2015; Hill, Hanton, Fleming, & Matthews, 2009; Hill, Hanton, Matthews, & Fleming, 2010), and in baseball (Gray, 2004; Otten & Barrett, 2013) have been shown to have lower than average performance in pressured situations. Conversely, studies have also shown evidence of success under pressure (e.g., see Jones, Hanton, & Connaughton, 2007; Turner & Barker, 2013), but these findings are less prevalent.

Sporting competition has been a common context for the study of choking under pressure because sports often present pressured situations and observable outcomes that

are widely available for analysis. As Sampaio, Lago, Casais and Leite (2010) note, sports presents an environment that involves psychological and physical demands that could lead to athletes incurring stress and reduced performance in pressured situations that can be difficult to find in non-sport related contexts. Beilock et al. (2004) noted that sports involve complex tasks that require attention in order to use working memory capacity, which can lead to possible distractions.

### **Basketball**

Mesagno et al. (2009) examined the influence of distraction on choking with experienced female basketball players. Players were grouped into four shooting groups, including: (a) low-pressure free throws without music; (b) high-pressure free throws without music; (c) low-pressure free throws with music; and (d) high-pressure free throws with music. The researchers found that participants improved their performance in high-pressure situations (with an audience) when there was music playing. Moreover, post-experiment surveys revealed that the music reduced the degree to which the participants were distracted.

Worthy et al. (2009) examined three seasons (2003–2004, 2004–2005 and 2005–2006) of National Basketball Association (NBA) free-throw percentages data. Binomial tests were used to determine if there were significant differences between free-throw shooting percentages during various game situations. When a team was down by two, down by one, or up by one, there were significant decrements in free-throw shooting near the end of the games as compared to near the beginning of games. Cao et al. (2011) also analyzed the effects of pressure on NBA free-throw data. Based on data from the 2002–2003 to 2009–2010 seasons, these scholars found that players shoot 5.5% lower than their

season average percentage when down by one and with less than 30 seconds remaining (significant at the 10% level) and 6.5% lower when trailing by one and with less than 15 seconds remaining (significant at the 5% level). Furthermore, Toma (2015) analyzed over 2.3 million NBA, Women's National Basketball Association (WNBA), and National College Athletic Association (NCAA) basketball free throws. With the exception of NCAA women, players' free-throw percentages decreased at the end of close games. This suggests that players are performing suboptimally during high pressure situations.

### **Golf**

Beilock (2007) conducted an experiment using 84 golf participants (42 with previous golf experience and 42 with no golf experience). The experienced golfers were high school or college players with a handicap of less than eight or with two or more years of varsity golf experience. The participants were split into four different groups based upon experience and whether they were using a regular putter or something considered a "funny" putter that had altered weight and shape. Participants undertook two blocks of 20 putts at first, followed by completing a memory questionnaire. An arithmetic test was then used to avoid potential recency issues. The dual tasks of putting and completing a questionnaire were then used as part of the experiment. Groups showed reduced putting skills in the dual task setting unless they were experienced golfers with the regular putter. Beilock (2007) had similar findings for those involved in the memory recognition test. These findings contrast explicit monitoring theory as it hypothesizes that attention to multiple details will lead to reduced performance in all cases. The experiment showed that too many tasks at once can be a distraction and lead to choking under pressure.



Hill et al. (2009, 2010) examined the performance of high-performance athletes in pressured situations, finding that athletes who focused on the outcome rather than the task itself had lower performance levels. In their first study, the researchers gathered qualitative data from four sport psychologists to determine that choking is a result of a stressful outcome that leads to a significant drop in performance. Hill et al. (2009) also argue that that individual differences and type of sport can have a role in whether individuals choke under pressure. Characteristics of an individual more likely to choke include low working memory capacity and lack of mental toughness. Hill et al. (2010), interviewed golfers and coaches of golfers who both had choked or had excelled under pressure. High expectations, thoughts of others, and unfamiliarity were some of the causes found for players to choke. Lowered self-confidence was a consistent outcome for golfers who had choked under pressure in the past.

To examine the impact of anxiety on success, Hickman and Metz (2015) analyzed 40,170 putts from 595 players in 353 tournaments to determine the impact of change to earnings on the likelihood that a Professional Golfers' Association (PGA) player would sink a putt on the final tournament hole. The likelihood of a player successfully sinking a putt decreased as the monetary value of a putt increased. The researchers estimated that every \$29,322 increase in prize money differential decreases the likelihood that the putt will be sunk by one percentage point.

## **Baseball**

Gray (2004) conducted baseball hitting experiments involving three groups that were: (a) just swinging, (b) swinging with a noise, and (c) swinging with a noise and saying if the bat was going up or down when the noise occurred. Experienced college

players with a mean of 13 years of experience were found to have their swing negatively impacted if they needed to explain if their swing was going up or down, whereas this would lead to increased performance by novices. In a second similar experiment, the factors were the same, but the players were tasked to respond to the noise on random occasions. For example, instead of responding every time, the response was given after the swing, not during, and only the expert hitters were used. No significant differences were found in terms of contact that was considered hits (based upon velocity and angle at which the ball made contact). In the third experiment, the factors were kept the same as the second experiment for 200 trials, and then participants were told they would be given \$20 if they increased their performance by 15% along with a teammate, having been told the teammate had already accomplished the goal. The results provide evidence of choking under pressure given the experimental design included an increased attentional focus to external items (prize and teammate performance) instead of to the task at hand. Athletes' overthinking then led to reduced performance.

### **Soccer**

To test whether pressured situations led to decreased performance, Dohmen (2008) compared home and visiting team penalty kick scoring rates in soccer. Based on a sample from the German Premier League of over 40 years of data including 3,619 penalty kicks, home teams were, on average, 2% less likely to score than players on visiting teams. Given the demanding nature of soccer fans, the researcher attributed the difference to pressure the crowd imposed on kickers. Further analysis showed the choking probability is highest when the score is tied (3.5% higher, significant at 5% level).

## **Tennis**

Williams and Rodrigues (2002) conducted an experiment that involved 10 amateur tennis players performing in varied working memory tasks under various levels of anxiety-inducing scenarios. In the low working memory task, players had to land the balls within one of three circles. In the high working memory test, the players had to land the ball within a specific large circle based upon where the server was positioned. In the low anxiety scenarios, players were told the results were only intended for the purposes of the study. In the high anxiety scenarios, they were advised that a \$200 prize would be awarded to the player that scored the most points. An ANOVA showed significant effects for anxiety in that the lower anxiety and working memory tasks led to significantly better results by the players.

## **Summary of Literature**

A variety of literature was found regarding choking under pressure. The primary finding is that the probability of success reduces when competitions are close in score, have limited time remaining, and are outside of the athlete's control. The literature in relation to sport was generally testing pressure filled situations that were individual specific. The researcher conducted an extensive search for literature that relates to scoreboard watching and nothing was found. Based upon general discussion of the phenomenon in sport (specifically baseball) the need for an academic study was apparent. The primary goal of the study is to determine if a generally discussed phenomenon exists and how it impacts performance. Another purpose is to fill a gap in choking under pressure research and how it relates to a collective.

## **CHAPTER IV**

### **EMPIRICAL APPROACH**

This chapter provides an overview of the research context, hypotheses, data, and empirical approach to this study. First, the context for analyzing the problem is described. Then, the data that will be under examination will be outlined. Third, specific hypotheses to be tested are laid out. The researcher will then explain how the data were collected and how the empirical model will be implemented.

#### **Context: Major League Baseball**

Game outcomes in Major League Baseball are analyzed in this study to determine if scoreboard watching occurs at the highest level of professional baseball. Each team plays 162 regular season games with most games (approximately 90%) played on consecutive days. This makes MLB an ideal context to examine scoreboard watching, as game information from the outcomes from the previous day may serve as a proxy variable. In addition, MLB had a change to its playoff structure in 2012 that further emphasized the value of being a division winner. This change meant that starting in 2012, only the division leaders qualified to play in the League Division Series, while the Wild Card teams now had a one-game playoff instead of directly qualifying for the Division Series, as an additional Wild Card team was added to each league. This implies that a team would be less satisfied with a Wild Card position as their season could be determined by one game, in which randomness could have a larger role in determining than team quality (i.e. not having the team's best pitcher available). Possible playoff revenue is also reduced by only being guaranteed one game.

The rules that qualified teams for postseason play from 2004 to 2014 were that the teams with the highest winning percentage from each of the six divisions (i.e., American League East, American League Central, American League West, National League East, National League Central, and National League West) would be in the playoffs each year. The playoff qualification rules changed in 2012. Each season from 2004 to 2011 had one Wild Card team in each of the American and National Leagues; this system had been in place since the 1995 season. The Wild Card teams were the best teams in each league in terms of winning percentages, but had not won a division. This meant there were a total of eight out of 30 teams qualifying for the playoffs. Each season since 2012 has had two Wild Card teams in each of the American and National Leagues. The Wild Card teams were the best two teams in each league in terms of winning percentages that did not win a division. These teams only qualified for a one-game, winner-take-all playoff, thus putting more emphasis on winning a division. This meant that since 2012 there has been a total of 10 out of 30 teams qualifying for the playoffs.

### **Hypotheses**

Based on the research questions, the following hypotheses have been formed:

1.  $H_{0A}$ : Scoreboard watching does not impact game outcomes.  
 $H_{1A}$ : Scoreboard watching has a negative impact on game outcomes.
2.  $H_{0B}$ : Games remaining has no impact on scoreboard watching.  
 $H_{1B}$ : There is a larger scoreboard watching effect with fewer games remaining.
3.  $H_{0C}$ : Scoreboard watching has the same influence on both the home and away teams.  
 $H_{1C}$ : Scoreboard watching influences home teams more than away teams.

4.  $H_{0D}$ : The addition of a Wild Card to each league had no impact on scoreboard watching.

$H_{1D}$ : Scoreboard watching has a larger negative effect with the addition of a new Wild Card to each league.

5. a)  $H_{0Ea}$ : Roster age has no impact on the magnitude of the scoreboard watching effect.

$H_{1Ea}$ : Rosters with older players are influenced by scoreboard watching less than rosters with younger players.

b)  $H_{0Eb}$ : Team payroll has no impact on the magnitude of the scoreboard watching effect.

$H_{1Eb}$ : Teams with higher payrolls are influenced by scoreboard watching less than rosters with lower payrolls.

c)  $H_{0Ec}$ : Managerial experience has no impact on the magnitude of scoreboard watching.

$H_{1Ec}$ : Teams with more experienced managers are influenced by scoreboard watching less than teams with little managerial experience.

d)  $H_{0Ed}$ : Momentum has no impact on the magnitude of the scoreboard watching effect.

$H_{1Ed}$ : Teams with momentum (i.e. winning streaks) are influenced by scoreboard watching less than teams without momentum.

### **Data Collection**

Game schedule and outcomes from the 2004 through the 2014 seasons, as well as each player's date of birth, salary, and the team manager's age were collected from

www.baseball-reference.com. These data also contained the game starting date and time, the identity of the home team and the team's division and conference. Furthermore, annualized team payroll data were collected from www.baseballprospectus.com.

Appendix A presents the code developed to identify the league and division leaders and to calculate the number of games back of each team, based upon their division. The final data set included 52,740 observations, where each observation included the identity of the division leader, whether the division leader won the previous day, the amount of games back of a team (from the division leader), the team's winning percentage, if a team played at home or played on the road, and the final score of each game.

### **Estimation Procedure**

These data were analyzed by testing for scoreboard watching with game level information that included: outcomes, competing team quality measures, and division leader previous game outcomes. Based on these factors, division leader results were excluded from the study. Specifically, the impact of scoreboard watching on game outcomes was expressed as:

$$(1) \text{Win}_{s,g,t} = \beta_0 + \beta_1 \text{scoreboard}_{s,g,t} + \beta_2 (\text{twinper}_{s,g,t} - \text{owinper}_{s,g,t}) + \varepsilon_{s,g,t}$$

The units of observation included: season ( $s$ ), game ( $g$ ), and team ( $t$ ). The win indicator variable was 1 if the team won and 0 if the team lost. The scoreboard indicator was 1 if the division leader won the previous day and 0 if they lost. The team winning percentage minus the opponent's winning percentage controlled for relative team quality. The  $\beta_1$  term directly tested for scoreboard watching. Specifically, the null hypothesis of no scoreboard watching is  $\beta_1 = 0$  and the alternative hypothesis is  $\beta_1 < 0$ . This is a method of

counterfactual observation, which determined an outcome based on if a variable was not present. In other models, home team indicator, roster age, manager's age, and team payroll variables are interacted with the scoreboard variable to determine their associated interaction effect. Equation (1) was estimated via Ordinary Least Squares estimate procedure.



## CHAPTER V

### RESULTS

In this chapter, the summary statistics and estimations results are presented. Multiple factors will be considered and presented throughout the tables. In this study, the coefficient of interest is the division leader win indicator variable as it presents the impact of scoreboard watching.

#### **Summary Analysis**

Performance of teams (measured by win/loss) relative to the prior result of the division leader was considered paramount for this study. In Table 1, game outcome summary statistics based on the 2004 to 2014 data are presented. This table shows how the unconditional probability of a win or loss is based upon if the division leader won or lost their previous game, as it demonstrates how well teams perform based on their knowledge of the result of the division leader's previous game. Team performance was examined on a per-month basis, following the hypothesis that teams will incur an increase in scoreboard watching toward the end of the season. This is because teams that are trailing the division leader will not only need to win their own games, they will also need the teams ahead of them to lose games in order to attain a playoff spot.

Given that a team has more games remaining to catch up to a division leader earlier in the season, more attention is given to optimizing performance than to considering external factors.

Another key variable is termed "the number of games back", which represents the amount of fewer wins a team has plus the amount of more losses a team has than the

Table 1: Team performance based on Division Leader's previous game outcome.

Month	Games Back	Division Leader's Previous Result	Team Winning Percentage	Standard Deviation
June	< 5	Loss	0.554	(0.497)
		Win	0.499	(0.500)
	5 to < 10	Loss	0.509	(0.500)
		Win	0.476	(0.500)
	$\geq 10$	Loss	0.478	(0.500)
		Win	0.444	(0.497)
July	< 5	Loss	0.582	(0.493)
		Win	0.513	(0.500)
	5 to < 10	Loss	0.563	(0.496)
		Win	0.428	(0.495)
	$\geq 10$	Loss	0.510	(0.500)
		Win	0.398	(0.490)
August	< 5	Loss	0.570	(0.495)
		Win	0.472	(0.499)
	5 to < 10	Loss	0.558	(0.497)
		Win	0.456	(0.498)
	$\geq 10$	Loss	0.508	(0.500)
		Win	0.419	(0.494)
September/October	< 5	Loss	0.641	(0.480)
		Win	0.506	(0.500)
	5 to < 10	Loss	0.625	(0.484)
		Win	0.454	(0.498)
	$\geq 10$	Loss	0.543	(0.498)
		Win	0.381	(0.486)

Results are based on all 2004–2014 MLB regular season game outcomes.

division leader. This number is then divided by two. For the purpose of this study the number of games back has been categorized into 1) less than five games back; 2) five games to less than 10 games back; and 3) greater than 10 games back. Categorizing in this way represents (a) a team being in the middle of a playoff race, (b) a team with an outside chance at qualifying for the postseason, and (c) a team with a very slim probability of qualifying for the playoffs. The hypothesis is that teams' winning percentages will be impacted by the division leader's result more when they are in the thick of a playoff race than teams that have close to a zero probability of securing a playoff position. The final explanatory variable in the table is if the division leading team won or lost their previous game. The hypothesis is that teams will put more pressure on themselves based on the external result if the team ahead of them in the standing wins instead of loses.

### **September and October**

Teams that are less than five games out of a playoff position have a winning percentage of 0.641 when the division leader loses their previous game during the months of *September* and *October*, and a winning percentage of 0.506 when the division leader wins. Teams that are five games to less than 10 games out of a playoff position have a winning percentage of 0.625 when the division leader loses their previous game during the months of *September* and *October*, and a winning percentage of 0.454 when the division leader wins. Teams that are 10 or more games out of a playoff position have a winning percentage of 0.543 when the division leader loses their previous game during the months of *September* and *October*, and a winning percentage of 0.381 when the division leader wins.

**August**

Teams that are less than five games out of a playoff position have a winning percentage of 0.570 when the division leader loses their previous game during the month of *August*, and a winning percentage of 0.472 when the division leader wins. Teams that are five games to less than 10 games out of a playoff position have a winning percentage of 0.558 when the division leader loses their previous game during the month of August, and a winning percentage of 0.456 when the division leader wins. Teams that are 10 or more games out of a playoff position have a winning percentage of 0.508 when the division leader loses their previous game during the month of August, and a winning percentage of 0.419 when the division leader wins.

**July**

Teams that are less than five games out of a playoff position have a winning percentage of 0.582 when the division leader loses their previous game during the month of *July*, and a winning percentage of 0.513 when the division leader wins. Teams that are five games to less than 10 games out of a playoff position have a winning percentage of 0.563 when the division leader loses their previous game during the month of July, and a winning percentage of 0.428 when the division leader wins. Teams that are 10 or more games out of a playoff position have a winning percentage of 0.510 when the division leader loses their previous game during the month of July, and a winning percentage of 0.398 when the division leader wins.

**June**

Teams that are less than five games out of a playoff position have a winning percentage of 0.554 when the division leader loses their previous game during the month

of *June*, and a winning percentage of 0.499 when the division leader wins. Teams that are five games to less than 10 games out of a playoff position have a winning percentage of 0.509 when the division leader loses their previous game during the month of June, and a winning percentage of 0.476 when the division leader wins. Teams that are 10 or more games out of a playoff position have a winning percentage of 0.478 when the division leader loses their previous game during the month of June, and a winning percentage of 0.444 when the division leader wins.

Based upon these findings, it can be stated that regardless of the month and number of games back, teams will have a higher winning percentage when the division leading team loses their previous game in comparison to having won their previous game. These results provide prefatory evidence that a relationship exists between the performance of division leading teams and the game result of other teams.

### **Estimation Results**

#### **Entire Season**

The impact of different variables on the probability of winning an MLB game are presented in Table 2. All regular season games from 2004 to 2014 are included, and divided into games from May to the end of the regular season (Model 1A); games from May to the end of August (Model 1B); games from September 1st to the end of the season (Model 1C); and only those games in which the team has not been eliminated from winning their given division (Models 1D and 1E). This latter category is characterized by having fewer games remaining than the difference in wins between the team and the division leader. Model 1B considers the part of the season that excludes the playoff race and therefore will have less expected value on scoreboard watching. Model

1C considers the part of the season that focuses exclusively on the playoff race, thus meaning that there would be an expected increase in the team's value on scoreboard watching. Model 1D considers the part of the season that focuses exclusively on the playoff race, but also excludes games where teams no longer have the potential to win their division. Model 1E consists of the same observations as Model 1D, but introduces the games back given the division leader won variable and the games back given the division leader lost variable. Addition of these variables is done to determine the interaction effect between scoreboard watching and how far a team is behind the division leader, to see if teams are likely to win based upon how far behind in the standings they are.

The dependent variable for Table 2 is the probability of winning a game. The variable of interest is whether the division leader won their previous game. The difference in winning percentage variable represents the competing team's winning percentage minus the winning percentage of their opposition. Therefore, the value will be negative if the team has a lower winning percentage than the opposition and will be positive if the winning percentage of the team is higher than the winning percentage of the opposition. The value will be zero if there is no difference between winning percentages of teams in a given game.

Model 1A, in which data from May to the end of each season from 2004 to 2014 are analyzed, has a constant of 0.555 and is significant at the 1% level. The division leader win indicator has a value of -0.097 and is significant at the 1% level. The difference in winning percentage variable has a value of 0.517 and is also significant at the 1% level. This can be summarized as a team having a 55.5% chance of winning a

game, given the division leader lost their previous game and there is no difference in the quality of two teams competing in a contest (based upon having the same winning percentage). A team is then 9.7% less likely to win a game if the division leader won their previous game. This would be a win probability of 45.8%. The difference in winning percentage variable can be described as a team being 5.17% more likely to win a game given they have a winning percentage that is 10% higher than their opposition. Therefore, if a team has a winning percentage of 55% and the opposition has a winning percentage of 45%, but the division leader won their previous game, the win probability would be 50.97% ( $55.5\% - 9.7\% + 5.17\%$ ).

Table 2: Ordinary Least Squares in relation to Scoreboard Watching and Win Probability

	Model 1A	Model 1B	Model 1C	Model 1D	Model 1E
Constant	0.555*** (0.004)	0.545*** (0.004)	0.591*** (0.009)	0.584*** (0.012)	0.618*** (0.024)
Division Leader Win Indicator	-0.097*** (0.005)	-0.081*** (0.006)	-0.158*** (0.011)	-0.147*** (0.016)	-0.155*** (0.030)
Difference in Winning Percentage	0.517*** (0.026)	0.468*** (0.029)	0.736*** (0.060)	0.874*** (0.097)	0.767*** (0.111)
Games Back Given Division Leader Won					-0.003 (0.002)
Games Back Given Division Leader Lost					-0.003* (0.002)
Number of Observations	35,386	27,915	7,471	3,839	3,839

Notes: \*\*\*, \*\*, \*, indicates significance at 1%, 5%, and 10% levels, respectively. Standard errors are in parentheses. Model 1A consists of all MLB regular season games from May to the end of the season for 2004 to 2014. Model 1B consists of all MLB regular season games May to the end of August for 2004 to 2014. Model 1C consists of all MLB regular season games from September 1st to the end of the season for 2004 to 2014. Models 1D and 1E consist of all MLB regular season games from September 1st to the end of the season for 2004 to 2014 in which the team has not been eliminated from winning their division.



Model 1B, in which data from May to the end of August for each season from 2004 to 2014 are analyzed, has a constant of 0.545 and is significant at the 1% level. The division leader win indicator has a value of -0.081 and is significant at the 1% level. The difference in winning percentage variable has a value of 0.468 and is also significant at the 1% level. This model can be interpreted as teams having a 54.5% chance of winning a game if the division leader lost their previous game and there is no difference in team quality. The division leader winning their previous game reduces the probability of a team winning a game to 46.4% (54.5% - 8.1%). The difference in winning percentage variable means a team will have a 4.68% higher chance of winning if the team's winning percentage is 10% higher than that of their opposition, or a 4.68% lower chance of winning if the opponent has a 10% higher winning percentage than the team.

Model 1C, in which data from September 1st to the end of each season from 2004 to 2014 are analyzed, has a constant of 0.591 and is significant at the 1% level. The division leader win indicator has a value of -0.158 and is significant at the 1% level. The difference in winning percentage variable has a value of 0.736 and is also significant at the 1% level. From these tests, this model can be interpreted as teams having a 59.1% chance of winning a game if the division leader lost their previous game and there is no difference in team quality. The division leader winning their previous game reduces the probability of a team winning a game to 43.3% (59.1% - 15.8%). The difference in winning percentage variable means a team will have a 7.36% higher chance of winning if the team's winning percentage is 10% higher than that of their opposition, or a 7.36% lower chance of winning if the opponent has a 10% higher winning percentage than the team.

Model 1D, in which data from September 1st to the end of each season from 2004 to 2014, and only games where the team has not been eliminated from winning their division, are analyzed, has a constant of 0.584 and is significant at the 1% level. The division leader win indicator has a value of -0.147 and is significant at the 1% level. The difference in winning percentage variable has a value of 0.874 and is also significant at the 1% level. This model can be interpreted as teams having a 58.4% chance of winning a game if the division leader lost their previous game and there is no difference in team quality. The division leader winning their previous game reduces the probability of a team winning a game to 43.7% (58.4% - 14.7%). The difference in winning percentage variable means a team will have a 8.74% higher chance of winning if the team's winning percentage is 10% higher than that of their opposition or a 8.74% lower chance of winning if the opponent has a 10% higher winning percentage than the team.

Model 1E has a constant of 0.618 and is significant at the 1% level. The division leader win indicator has a value of -0.155 and is significant at the 1% level. The difference in winning percentage variable has a value of 0.767 and is also significant at the 1% level. The games back given the division leader won variable has a value of -0.003 and is not significant. The games back given the division leader lost variable has a value of -0.003 and is significant at the 10% level. This can be interpreted as a team having a 61.8% chance of winning a game given the division leader lost the previous game, the opponent is of equal quality and the team is now tied with the division leader in the standings. Winning percentage then decreases by 15.5% if the division leader won their prior game. If the difference in winning percentage between the team and the opposition is 10% higher for the team of interest, the probability of winning a game will

increase by 7.67%. For every additional game back a team is from the division leader, their win probability will decrease by 0.3%, regardless of whether the division leader won or lost their previous game.

The division leader win indicator is consistently negative, showing that teams are negatively impacted by scoreboard watching. The absolute value of this variable increases from 0.081 to 0.158, an increase of 7.7%, when looking at the early season to the end of August in comparison to regular season games in September and October. A 7.7% difference in probability of winning based on an external outcome seems extreme and is another reason teams should consider eliminating scoreboard watching. Only considering games where teams have not been eliminated from winning their division in comparison to all September/October regular season games produced similar results. This suggests that teams are likely to perform similarly during the playoff race whether they have been eliminated from winning their division or not, possibly due to having an opponent that is still in the race.

### **Playoff Races**

Table 3 presents the results based on teams that are in the playoff race. A team in a playoff race is determined by being five or fewer games behind the division leader and not eliminated from the playoffs. Columns 1 and 3 include August games from non-division leaders in which teams are “in the playoff race”. Column 1 consists of only three variables (Constant, Division Leader Win Indicator, and Difference in Winning Percentage). Column 3 consists of these variables as well as the interaction effects of Games Back Given Division Leader Won and Games Back Given Division Leader Lost. Column 2 has the same characteristics as column 1, but is regular season games from

September to the end of the regular season. Column 4 has the same characteristics as column 3, but also represents regular season games from September to the end of the regular season.

For August games of non-division leaders in which teams are in the playoff race, the value of the constant is 0.547 and is significant at the 1% level. The division leader win indicator has a value of -0.096 and also has significance at the 1% level. Difference in winning percentage is significant at the 1% level and has a coefficient of 0.499. This shows that teams are 9.6% less likely to win a game if the division leader won the previous day in August. The first September column has a constant of 0.600 and is significant at the 1% level. The division leader win indicator for September has a value of -0.121 and also has significance at the 1% level. Difference in winning percentage is significant at the 5% level and has a coefficient of 0.526. This shows that teams are 12.1% less likely to win a game if the division leader won the previous day during the months of September and October. This shows that teams are 2.5% less likely to win a game in September/October than August if the division leader won their previous game.

The second August column has a constant of 0.489 and is significant at the 1% level. The division leader win indicator has a value of -0.013, but is not significant. Difference in winning percentage is significant at the 1% level and has a coefficient of 0.513. Games back given a division leader won has a coefficient of -0.008 and is not significant. Games back given a division leader lost has a coefficient of 0.019 and is not significant. This shows that teams are 1.3% less likely to win a game if the division leader won the previous day in August. The second September column has a constant of 0.611 and is significant at the 1% level. The division leader win indicator for September

has a value of -0.068, but lacks significance. Difference in winning percentage is significant at the 5% level and has a coefficient of 0.495. Games back given a division leader won has a coefficient of -0.021 and is not significant. Games back given a division leader lost has a coefficient of -0.003 and is not significant. This shows that teams are 6.8% less likely to win a game if the division leader won the previous day during the months of September and October. This shows that teams are 5.5% less likely to win a game in

Table 3: Ordinary Least Squares in relation to Scoreboard Watching and Win Probability for Teams in Race

	August	Sept/Oct	August	Sept/Oct
Constant	0.547*** (0.021)	0.600*** (0.026)	0.489*** (0.048)	0.611*** (0.053)
Division Leader Win Indicator	-0.096*** (0.025)	-0.121*** (0.031)	-0.013 (0.059)	-0.068 (0.070)
Difference in Winning Percentage	0.499*** (0.164)	0.526** (0.227)	0.513*** (0.168)	0.495** (0.229)
Games Back Given Division Leader Won			-0.008 (0.011)	-0.021 (0.014)
Games Back Given Division Leader Lost			0.019 (0.014)	-0.003 (0.016)
Number of Observations	1,590	995	1,590	995

Notes: \*\*\*, \*\*, \*, indicates significance at 1%, 5%, and 10% levels, respectively. Standard errors are in parentheses. Teams in the race are those that are five games back or fewer and have not been eliminated from winning division.

September/October than August if the division leader won their previous game. After adding the interaction effect, the scoreboard watching indicator variable of if the division leader won or lost is no longer significant. This could be due to the fact that there is a strong correlation between the interaction effects and the variable of interest of scoreboard watching. As the time moved from August to September, all coefficients relating to scoreboard watching (Division Leader Win Indicator, Games Back Given Division Leader Won, and Games Back Given Division Leader Lost) saw their values become higher negative numbers, supporting the notion that scoreboard watching is a significant phenomenon.

### **Team Fixed Effects**

In Table 4, the results from a team fixed effects regression are presented. This model accounts for potential team effects beyond within-season winning percentages. Observations included are the same as in Table 3; that is, games in September where teams have not been eliminated from winning their division and are five or fewer back from the division leader, from 2004 to 2014. The Washington Nationals coefficient also relates to the 2004 Montreal Expos because the latter team moved to Washington for the beginning of the 2005 season. Through running this regression, it was determined that there were no results for the Chicago Cubs, Seattle Mariners, or Toronto Blue Jays. This shows that these teams were either leading their division or more than five games back from the division leader at all points from September to the end of the regular season for any given season. This is justifiable as neither the Toronto Blue Jays nor Seattle Mariners qualified for the postseason from 2004 to 2014. In order to avoid endogeneity issues, the Arizona Diamondbacks were excluded from the regression.

The division leader win indicator is  $-0.117$  and is significant at the 1% level. Difference in winning percentage is significant at the 1% level and has a value of  $0.492$ . Each team then has a coefficient, ranging from  $-0.382$  for the Miami Marlins to  $0.314$  for the Washington Nationals. The coefficients for all of the teams are not significant. This can be interpreted as teams' winning percentage probabilities decreasing from  $63.6\%$  to  $51.9\%$  if the division leader won their previous game as compared to them having lost it. Given a division leader lost the previous day, the winning percentage of the teams competing in a matchup are  $13.7\%$  for the Miami Marlins and  $83.3\%$  for the Washington Nationals, with all other teams having winning percentages falling in between this range. There could be a lack of observations (995), leading to possibly skewed results for team fixed effects. However, the results point to teams like the Washington Nationals ( $0.314$ ), Pittsburgh Pirates ( $0.141$ ), Baltimore Orioles ( $0.122$ ), and New York Mets ( $0.110$ ) having the proper procedures in place to be successful during a playoff race. Each team is 10% more likely than the average team to win a game during the playoff race. The key finding is that specific team results do not take away from the overall/general significance of the scoreboard watching variable (Division Leader Win Indicator).



Table 4: Team Fixed Effects Model in relation to Scoreboard Watching and Win Probability

	Sept/Oct		Sept/Oct
Constant	0.636*** (0.110)	Miami Marlins	-0.382 (0.245)
Division Leader Win Indicator	-0.117*** (0.032)	Milwaukee Brewers	-0.158 (0.132)
Difference in Winning Percentage	0.492** (0.239)	Minnesota Twins	0.044 (0.123)
Atlanta Braves	-0.115 (0.140)	New York Mets	0.110 (0.189)
Baltimore Orioles	0.122 (0.157)	New York Yankees	0.028 (0.131)
Boston Red Sox	-0.064 (0.122)	Oakland Athletics	-0.048 (0.126)
Chicago White Sox	-0.176 (0.138)	Philadelphia Phillies	-0.062 (0.125)
Cincinnati Reds	-0.031 (0.143)	Pittsburgh Pirates	0.141 (0.135)
Cleveland Indians	-0.055 (0.140)	San Diego Padres	-0.014 (0.126)
Colorado Rockies	0.033 (0.128)	San Francisco Giants	-0.061 (0.120)
Detroit Tigers	-0.021 (0.138)	St. Louis Cardinals	-0.158 (0.184)
Houston Astros	0.069 (0.196)	Tampa Bay Rays	-0.057 (0.143)
Kansas City Royals	-0.058 (0.167)	Texas Rangers	-0.123 (0.133)
Los Angeles Angels	-0.054 (0.126)	Washington Nationals	0.314 (0.505)
Los Angeles Dodgers	-0.011 (0.135)	Number of Observations	995

Notes: \*\*\*, \*\*, \*, indicates significance at 1%, 5%, and 10% levels, respectively. Standard errors are in parentheses. Table 4 consists of all MLB regular season games in September and October for 2004–2014 in which teams are five games or less back from the division leader and not eliminated from winning their division. Please note that the Washington Nationals also represent the 2004 Montreal Expos who relocated to Washington, DC, in 2005. Arizona Diamondbacks were excluded from the Fixed Effects model to avoid endogeneity issues. The Chicago Cubs, Seattle Mariners, and Toronto Blue Jays did not meet the parameters.

## Game Location

To determine whether scoreboard watching influences teams playing at home more than teams playing away from home, the estimation results based on home and away team specific models during the September/October playoff race are presented in Table 5. The division leader win indicator is -0.233 and significant at the 1% level in the home specific model, and -0.009 and not significant in the away team model. Teams initially have a higher probability of winning at home (0.667 compared to 0.528 for away teams), which suggests that teams playing at home have an advantage. The finding suggests that teams are putting additional pressure on themselves to win in front of their home crowd based upon the outcome of the division leader. As such, evidence of an increased probability of choking under pressure for teams playing at home is consistent with distraction theory.

Table 5: Ordinary Least Squares in relation to Scoreboard Watching and Win Probability for Teams in Race based upon Game Location

	Home	Away
Constant	0.667*** (0.036)	0.528*** (0.038)
Division Leader Win Indicator	-0.233*** (0.044)	-0.009 (0.045)
Difference in Winning Percentage	0.409 (0.321)	0.553* (0.320)
Number of Observations	487	508

Notes: \*\*\*, \*\*, \*, indicates significance at 1%, 5%, and 10% levels, respectively. Standard errors are in parentheses. Teams in the race are those that are five games back or fewer and have not been eliminated from winning their division.

## **Wild Card**

In Table 6 the results based on games played between 2004 to 2011 and 2012 to 2014 are presented. As discussed in Chapter IV, additional Wild Card teams were introduced to both the American and National Leagues in 2012. This alteration to league structure added additional value for teams to win their division, thus leading to the hypothesis that scoreboard watching of the division leader by teams in the playoff race would become a stronger factor. Starting in 2012, being a Wild Card team only guaranteed a one-game playoff in comparison to a Best of 5 series. Having only one game with a random result is expected to lead to more emphasis on winning the division.

The 2004–2011 division leader win indicator has a value of -0.113 and is significant at the 1% level, while the 2012–2014 division leader win indicator has a value of -0.152 and is significant at the 5% level. Teams are therefore 3.9% less likely to win a game in a playoff race if the division leader won their previous game for the years 2012 to 2014, based upon a division leader win indicator going from -0.113 to -0.152. This is a sign that scoreboard watching could be more common with the current league structure.

Table 6: Ordinary Least Squares in relation to Scoreboard Watching and Win Probability for Teams in Race based upon Season

	2004–2011	2012–2014
Constant	0.604*** (0.030)	0.587*** (0.051)
Division Leader Win Indicator	-0.113*** (0.036)	-0.152** (0.062)
Difference in Winning Percentage	0.316 (0.255)	1.416*** (0.506)
Number of Observations	741	254

Notes: \*\*\*, \*\*, \*, indicates significance at 1%, 5%, and 10% levels, respectively. Standard errors are in parentheses. Teams in the race are those that are five games back or fewer and have not been eliminated from winning their division. Additional Wild Card team added to each league in 2012.

### Games Remaining

In Table 7, the impact on a team of amount of games remaining is presented, divided in the following ways: more than 20 games remaining during a playoff race in September/October (column 1), greater than 10 to 20 games remaining (column 2), and 10 or fewer games remaining (column 3). The purpose of these data in this table is to determine if games at the very end of the regular season are treated differently than games at the beginning of the defined playoff race. The hypothesis is that with fewer games remaining, teams are more likely to scoreboard watch as players/teams have less time to catch up to the division leader and teams need both the leader to lose and they themselves to win to take the division lead.

Based on greater than 20 games remaining, and greater than 10 games remaining to 20 games, the division leader win indicator variables are not significantly different from zero. However, with less than 10 games remaining, the division leader win indicator has a value of -0.243 with significance at the 1% level. As fewer games are remaining,

the absolute value of the division leader win indicator goes up, supporting the notion that teams engage in scoreboard watching more toward the season's end.

Table 7: Ordinary Least Squares in relation to Scoreboard Watching and Win Probability for Teams in Race based upon Games Remaining

	> 20 GR	>10 to 20 GR	≤ 10 GR
Constant	0.631*** (0.048)	0.537*** (0.043)	0.646*** (0.044)
Division Leader Win Indicator	-0.053 (0.058)	-0.068 (0.051)	-0.243*** (0.055)
Difference in Winning Percentage	-0.438 (0.405)	0.909** (0.377)	1.130*** (0.401)
Number of Observations	303	379	313

Notes: \*\*\*, \*\*, \*, indicates significance at 1%, 5%, and 10% levels, respectively. Standard errors are in parentheses. Teams in the race are those that are five games back or fewer and have not been eliminated from winning their division.

### Roster Age

In Table 8, the impact of average age of the roster is examined. This is further analyzed by considering pitcher and batter ages. The purpose of this table is to see if teams should look for older or younger players to properly deal with scoreboard watching based upon performance of age groups. The hypothesis is that teams with older players will be more likely to win as these players may have experienced past playoff races and thus will be more likely to give advice of what to do daily to avoid the negative impacts of scoreboard watching.

The average age of a team was found to be not significant in predicting the outcome of a game and the value of the scoreboard watching variable remains significant at the 1% level with a reduced win probability of 0.123. Moreover, considering roster positions (batter and pitcher) did not change the significance of the scoreboard watching

variable. Of note is that the coefficient for batters is negative and the coefficient for pitchers is positive. These values are not significant so they are not strong predictors of game outcomes.

Table 8: Ordinary Least Squares in relation to Scoreboard Watching and Win Probability for Teams in Race in September based upon Average Age

	Team Average	Position Average
Constant	0.601*** (0.026)	0.602*** (0.026)
Division Leader Win Indicator	-0.123*** (0.032)	-0.126*** (0.032)
Difference in Winning Percentage	0.514** (0.231)	0.492** (0.232)
Log of Difference in Average Age Interacted with Division Leader Outcome	0.122 (0.386)	
Log of Difference in Average Age Interacted with Division Leader Outcome (Batters)		-0.172 (0.384)
Log of Difference in Average Age Interacted with Division Leader Outcome (Pitchers)		0.325 (0.342)
Number of Observations	995	995

Notes: \*\*\*, \*\*, \*, indicates significance at 1%, 5%, and 10% levels, respectively. Standard errors are in parentheses. Teams in the race are those that are five games back or fewer and have not been eliminated from winning their division.

## **Team Payroll**

In Table 9, the impact of team payroll is examined, wherein data included consists of both total dollar payroll and team ranking for payroll. The purpose of this table is to examine whether the amount of money a team spends impacts the probability of winning a game. The hypothesis is that teams that spend more money on players will be more likely to win as they are more likely to have high quality players who consistently exhibit high performance metrics.



Table 9: Ordinary Least Squares in relation to Scoreboard Watching and Win Probability for Teams in Race in September based upon Team Payroll

	Team Payroll (\$)	Team Payroll (Rank)
Constant	0.602*** (0.026)	0.602*** (0.026)
Division Leader Win Indicator	-0.123*** (0.032)	-0.124*** (0.032)
Difference in Winning Percentage	0.505** (0.233)	0.489** (0.233)
Log of Difference in Team Payroll Interacted with Division Leader Outcome	0.016 (0.038)	
Log of Difference in Team Payroll Rank Interacted with Division Leader Outcome		-0.013 (0.017)
Number of Observations	995	995

Notes: \*\*\*, \*\*, \*, indicates significance at 1%, 5%, and 10% levels, respectively. Standard errors are in parentheses. Teams in the race are those that are five games back or fewer and have not been eliminated from winning their division.

After considering team payroll and rank, the significance of the division leader win indicator remains unchanged ( $p < 0.01$ ) and the absolute value of the coefficient remains at 0.123 (difference in \$ payroll) and 0.124 (difference in rank of payroll). The team with the higher payroll has a slight increase in probability of winning, but not to a level in which payroll is a significant factor.

### **Managerial Experience**

In Table 10, the degree to whether manager's experience can influence the probability of winning a game is exhibited. The first column presents the estimation results based on managerial playoff experience, the second column based on the number of season managed, and the third column based on the percentage of seasons that the manager qualified for the playoffs.

The purpose of this table is to see if a manager's experience can influence the probability of winning a game. The hypothesis is that teams that hire a manager who has accrued more playoff experience will be more likely to win, as they have experienced playoff races in the past and were able to successfully qualify for the postseason before. Thus, the expectation is that the manager's style may be a contributing factor in qualifying for the playoffs.

Table 10: Ordinary Least Squares in relation to Scoreboard Watching and Win Probability for Teams in Race in September based upon Managerial Statistics

	Playoff Appearances	Seasons Managed	% of Seasons in Playoffs
Constant	0.598*** (0.026)	0.597*** (0.026)	0.597*** (0.026)
Division Leader Win Indicator	-0.120*** (0.032)	-0.121*** (0.032)	-0.121*** (0.032)
Difference in Winning Percentage	0.500** (0.232)	0.506** (0.230)	0.502** (0.233)
Manager Playoff Appearances Interacted with Division Leader Outcome	0.004 (0.004)		
Manager Seasons Managed Interacted with Division Leader Outcome		0.002 (0.002)	
Manager's % of Playoff Appearances Interacted with Division Leader Outcome			0.053 (0.067)
Number of Observations	995	995	995

Notes: \*\*\*, \*\*, \*, indicates significance at 1%, 5%, and 10% levels, respectively. Standard errors are in parentheses. Teams in the race are those that are five games back or fewer and have not been eliminated from winning their division.

The number of seasons managed and playoff appearances are not significant factors in explaining game outcomes. However, the coefficients for scoreboard watching remain significant at the 1% level and range from 0.120 and 0.121. Teams with managers who have accrued more managerial experience and seasons in the playoffs have only a slight increase in win probability with each additional season managed, in comparison to the opposition, thereby increasing the win probability by 0.004 and each additional playoff experience compared to the opposition increasing win probability by 0.002.

### **Momentum**

In Table 11, the impact of team momentum is examined.. The first of three columns depicts the difference in winning/losing streaks among two teams in a given matchup, where winning streaks are positive and losing streaks are negative. Column 2 presents the teams in the playoff race that are on winning streaks. Column 3 presents teams that are on a losing streak when in playoff races. A winning or losing streak could be for only one game for this study. For the purpose of this research a streak is amount of games a team has consecutively won or lost. If teams had an equal streak going into a matchup then the difference in the streak after the game would be two as one team would be positive 1 and the other would be negative 1. The purpose of this table is to see if momentum impacts the probability of winning a game. The hypothesis is that teams that are on a streak of consecutive win/loss outcomes are more likely to continue the streak if momentum is an existing phenomenon. Therefore, winning streaks would have a positive coefficient and losing streaks would also have a positive coefficient.

Table 11: Ordinary Least Squares in relation to Scoreboard Watching and Win Probability for Teams in Race in September based upon Winning/Losing Streak

	Momentum	Given Winning Streak	Given Losing Streak
Constant	0.599*** (0.026)	0.637*** (0.033)	0.540*** (0.043)
Division Leader Win Indicator	-0.117*** (0.032)	-0.147*** (0.053)	-0.121* (0.062)
Difference in Winning Percentage	0.548** (0.229)	0.482* (0.283)	0.629** (0.389)
Difference in Streak Interacted with Division Leader Outcome	-0.004 (0.005)	-0.002 (0.009)	-0.020 (0.013)
Number of Observations	995	602	393

Notes: \*\*\*, \*\*, \*, indicates significance at 1%, 5%, and 10% levels, respectively. Standard errors are in parentheses. Teams in the race are those that are five games back or fewer and have not been eliminated from winning their division.

Difference of streak between two teams in a given matchup and teams on winning streaks continue to have a negative coefficient for the scoreboard watching variable that is significant at the 1% level. However, teams on losing streaks see the division leader win indicator variable's significance only at the 10% level. The coefficient remains similar to the other models (-0.121). Difference in streak (regardless of if a team is on a winning or losing streak) does not have significant values. Of note is that these values are all negative. This suggests that the team competing with the more positive streak is less likely to win. This is specifically shown in that the coefficient for teams on losing streaks is the highest, suggesting that the longer the losing streak a team incurs, the more likely they are to end the streak and win the next game.

### **Opposition**

In Table 12, the impact of the opposition is measured. The opposition being the division leader is presented in Column 1. The performance against any team that is not leading a division (excludes AL and NL East, Central and West Leaders) is considered in Column 2. With regards to Column 3, results are presented in terms of games against teams that are leading a division, but not the division that the team being analyzed is in. The purpose of this table is to determine if any of the scoreboard watching behaviour is determined by head to head matchups with division leaders. Based upon MLB being a league of primarily three game series, that would mean that two games in the series the team would have direct control for if the division leader lost. The expectation is that regardless of the opposition, the scoreboard watching phenomenon will lead to a reduced win probability if the division leader won the previous day.

Table 12: Ordinary Least Squares in relation to Scoreboard Watching and Win Probability for Teams in Race based upon Opponent Ranking

	Division Leader	Non-Division Leader	Other Division Leader
Constant	0.938*** (0.027)	0.532*** (0.034)	0.328*** (0.111)
Division Leader Win Indicator	-0.967*** (0.023)	0.023 (0.036)	0.171 (0.140)
Difference in Winning Percentage	-1.371 (1.177)	0.418 (0.280)	-2.282 (2.075)
Number of Observations	153	791	51

Notes: \*\*\*, \*\*, \*, indicates significance at 1%, 5%, and 10% levels, respectively. Standard errors are in parentheses. Teams in the race are those that are 5 games back or fewer and have not been eliminated from winning their division.

The findings in Table 12 suggest that scoreboard watching is only negative and significant at the 1% level when a team is facing their division leader. This outcome appears reasonable as if the division leader lose the day before; the high likelihood (based on MLB's series structure) is that the team being analyzed is the team that competed against the division leader. The value's for the coefficients being approximately 1 for the constant and -1 for the division leader win indicator are rational as the team/division leader will win and the other will lose as there are no ties in MLB. When playing against non-division leaders, the result of the division leader the previous day has very little impact on the game outcome.



## CHAPTER VI

### DISCUSSION

This chapter discusses the implications of the results found for scoreboard watching. The discussion first revisits the hypotheses that were presented in Chapter IV. Additional findings of interest from the analysis will then be considered. The chapter will then describe the limitations to this study. Suggestions for potential future studies that can further investigate if a scoreboard watching phenomenon exists will conclude this chapter.

#### **General Findings**

Division leader wins had a significantly negative impact on a team's probability of winning a game. Multiple factors were controlled, and regardless of the factor, the coefficient of interest for scoreboard watching was negative. This shows that the findings are robust. The factors that influenced the scoreboard watching variable the most were if the game was played at home or on the road (0.233 reduced probability of winning at home if division leader won previous day,  $p < 0.01$ , as seen in Table 4), the amount of games remaining in the season (0.243 reduced probability of winning with 10 or fewer games remaining in a season if division leader won the previous day,  $p < 0.001$ , as seen in Table 7), and if it was during the beginning or end of the season (0.158 reduced probability of winning in September if division leader won the previous day,  $p < 0.001$ , as seen in Table 2). Evidence of reduced probability of winning based on scoreboard watching can be found throughout the Results section. Therefore,  $H_{0A}$  (scoreboard watching does not impact game outcomes) can be rejected.

The null hypothesis ( $H_{0B}$ ) that scoreboard watching is not impacted by games remaining can be rejected based upon Table 2. Teams are less likely to win by 8.1% based upon scoreboard watching from May to the end of August. This increases to 15.8% in the months of September and October. This could represent that teams are paying more attention to what is being presented in the media at later points of the season. This is possibly a sign that teams are choking under pressure later in the season, possibly due to their belief that there is greater value to games being played at the end of the season.

When there are fewer games remaining in the season, there appears to be a larger influence of scoreboard watching. Specifically, this becomes more noticeable when there are 10 or fewer games remaining in a season. In Table 7, there is only approximately a 6% chance that teams are more likely to lose given a division leader won the previous game (5.3% if 20 or more games remaining and 6.8% if greater than 10 to 20 games remaining). This increases significantly if there are 10 or fewer games remaining. The coefficient for scoreboard watching is then 0.243. If a division leader wins six of their final 10 games, this would then lead to the team behind them losing an additional 1.5 games on average. So, even if this is only one game in a given case, this is still a meaningful factor. An example is that the New York Mets (88–74) of 2007 missed the playoffs by one game. Their division winner was the Philadelphia Phillies (89–73). With an additional win, the Mets would have had a tiebreaker game to qualify for the postseason, and if they won that game, then several other outcomes may have been affected.

In terms of home teams, if a team was considered to be in the playoff race and had 16 home games remaining on their schedule, and if the team that was leading the division won their prior game the day before half the time (eight games), based on the coefficient of 0.233 for scoreboard watching given in Table 5, the team would then lose two additional games. This could be the difference between winning a division and not qualifying for postseason play. If this is the case, it would be preferable for a team to play on the road as the scoreboard watching coefficient is close to zero (-0.009). This could be a sign that teams are becoming too comfortable in their settings, allowing a possible overconfidence effect of playing on home field to take place. Based on these details,  $H_{0C}$  can be rejected.

Possible strategies to minimize or eliminate this issue would be to treat home stands like road trips and stay together in a team hotel. Teams could take away some of the additional luxuries provided in their home clubhouse as well. If the team is being given additional information from their staff, this could be a possible distraction, so this could also be eliminated. The schedule is out of the team's control, so any additional ways to treat home games more like road games would be useful. If scoreboard watching is seen to be an issue, then the team could look into the possibility of eliminating the live updates to scores on the fences. Fans could continue to access this information themselves through cellular phones and using wireless internet supplied by most stadiums.

The research design allowed for the current playoff structure (two Wild Card teams per league) to be compared to the prior playoff structure (one Wild Card team per league). Results in Table 6 show teams being 3.9% less likely to win a game with the

new structure (2012–2014 data) than with the prior (2004–2011 data). As this table represented September/October games, if the division leader had a record of 18–12 during this period, the team behind them in the divisional playoff race would see their number of wins decrease by 0.702. If this value is rounded up to one, this could be the difference between making the playoffs or not (as shown in the previously described case of teams playing at home as compared to away). This supports the notion that teams are misjudging the impact of the additional Wild Card and added value to being a divisional winner. This further suggests that league/playoff structure is an external factor that is being over weighted in relation to scoreboard watching. This provides sufficient evidence to reject  $H_{0D}$ .

Based upon Table 8,  $H_{0Ea}$  (roster age has no impact on the magnitude of the scoreboard watching effect) cannot be rejected. Whether considering average age of the entire roster, or separating it into pitchers and batters, the results for the variable were not significant. This means that there is not a specific need for the general manager to target younger or older players when in a playoff race. An interesting finding is that the coefficient is negative for batters and positive for pitchers. This is a sign that younger batters and older pitchers are possibly better able to handle playoff races. This could be a sign that pitchers are better with experience and could use their experience to strike a batter out in a difficult situation. The absolute value for pitchers is also higher for pitchers than for batters. This could mean that a top priority for general managers at the trade deadline could be to bring in a veteran pitcher.

With team payroll under consideration,  $H_{0Eb}$  cannot be rejected. This hypothesis suggests that having a higher team payroll will not make you significantly more likely to

win during a playoff race. The coefficient for the variable is positive in terms of dollars so it suggests that a team can in fact enhance their win probability by increasing their payroll. Adjusting payroll, however, is not an easy change to make. It is dependent on allocations that the team ownership makes. If a team is lacking in ticket and merchandise sales then increasing team payroll is unlikely. However, player salary is correlated to player quality, so if a reasonable trade is a possibility at the trade deadline and not too high of a value of prospects will be given up, a suggestion is to make a trade to take on additional salary.

Managerial experience was considered in Table 10. In terms of seasons managed, playoff appearances, and percentage of seasons managed being in the playoffs, the results were not significant.  $H_{0Ec}$  cannot be rejected due to this. Each of the coefficients was above zero, so this would infer that hiring a manager with lots of regular season and playoff experience would be ideal. The issue with this is that it can be expected that all teams would want a manager with those credentials. There are only limited managers who fit these criteria and most are already under contract by teams, while others could have been through a recent streak of missing the playoffs. However, hiring members of that manager's coaching staff could be an effective solution for gaining an affordable manager who has the benefit of having learned from another successful manager.

Information in relation to the effects of momentum was presented in Table 11. A better winning streak than the opposition consistently led to a negative coefficient regarding the probability of winning a game. This means a longer winning streak, a shorter losing streak, or a winning streak when the opposition had a losing streak reduced the probability of winning a game. However, the results were not significant, so  $H_{0Ed}$

cannot be rejected. This supports the notion that momentum is not a significant phenomenon and that team quality is more of a determinant in game outcomes. An interesting finding is that the bigger the difference in the streak, the more likely the team with a negative streak is to win a game. So, given a team is on a losing streak of seven games and the opposition is on a five-game winning streak, the probability of the team on the losing streak winning a game will increase by 24%  $((-7-5)*-2\%)$  in comparison to both teams being on a one-game losing streak.

### **Additional Findings**

A quick glance at the results of this study, particularly the finding that the scoreboard watching indicator variable has a consistently negative value, one could get the impression that a team in a playoff race is not going to qualify for the playoffs at any point. This is not the case. If the division leader were to go on a losing streak or even have a record just below .500, there are many opportunities for a team to surpass them in the standings. If a division leader were to lose the previous game, the only two variables in the regression equation would be the constant and the difference in team quality in a game. As the constant is consistently above 0.5, and the team faces a schedule of teams with lower winning percentages than them, this is the situation in which a team is most likely to come from behind in the standings to win a division. So, the existence of scoreboard watching and one team seeing another falter in front of them could actually be seen as positive, as teams are motivated to make up ground. They could take a “blood in the water” approach, where seeing another team struggling leads to their own momentum to win. This in turn could lead to the division leader looking at the standings and becoming anxious of the team behind them in the standings.

Reviewing the data from 2004 to 2014 (11 seasons), teams tied for the division lead twice, finished one game back of the division lead 11 times, finished two games back of the division lead 10 times, and finished three games back of the division lead on six occasions. This is a total of 29 times missing the division title by three games or less. That is almost three teams a year. Several of these cases involved the teams qualifying for the playoffs through a Wild Card position, but there are several more that missed the playoffs altogether. The results show that scoreboard watching can make the difference of just a few games, and that teams could have had home field advantage or made the playoffs by not placing their focus on the performance of other teams against which they are not matched up.

This analysis adds to the literature reviewed in Chapter IV in many aspects. Choking under pressure (specifically distraction theory) was applied throughout the analysis and consistently had a significant negative impact on win probability. Home field advantage had a significant effect on winning as teams were impacted more by scoreboard watching at home than away, and playoff/league structuring was an evident factor as the playoff structure change in 2012 led to a larger effect of scoreboard watching. While the analysis of momentum, coaching/leadership, and team composition had no significant effects, these factors for analysis still provided valuable results.

### Hypothesis test summary

Null hypothesis	Test result	Magnitude of the effect
Scoreboard watching does not impact game outcomes.	Reject	-0.097
Games remaining has no impact on scoreboard watching.	Reject	-0.243 with 10 or fewer games remaining
Scoreboard watching has the same influence on both the home and away teams.	Reject	-0.233 for home teams and -0.009 for away teams
The addition of a Wild Card to each league had no impact on scoreboard watching.	Reject	Went from -0.113 (2004-11) to -0.152 (2012-14)
Roster age has no impact on the magnitude of the scoreboard watching effect.	Do not reject	-0.123
Team payroll has no impact on the magnitude of the scoreboard watching effect.	Do not reject	-0.123
Managerial experience has no impact on the magnitude of the scoreboard watching effect.	Do not reject	-0.120
Momentum has no impact on the magnitude of the scoreboard watching effect.	Do not reject	-0.117

### Limitations

Through regression analysis, scoreboard watching is found to be a significantly negative influence on teams in MLB. Multiple tables applying OLS showed the indicator variable representing scoreboard watching to be significant consistently at the 5% level. However, there are limitations to this study. A limitation of the indicator variable is that it focuses only on the division leader's game outcome from the previous day. This does not take into account that the division leader could have had a game earlier on the same day, as baseball games are played at a variety of times, especially on weekends. There are also



different time zones that teams play within, so a game in the Eastern Time zone for the division leader has a high likelihood of being completed before a game in the Pacific Time zone. This would mean that there could be more recent game outcomes that teams are viewing before their games.

Another possible limitation is that other control variables could be used for the regression equation. Winning percentage is an effective measure of team quality, but other factors could be added. These statistics include pregame Earned Run Average of the starting pitchers, the specific starting lineup for the team and their batting averages, and the Earned Run Average of the bullpen. The issue is that these factors are not readily available. Box scores contain the information after the completion of the team's game. This study requires the pregame information and would then require finding the result of each player's previous game as lineups in Major League Baseball are consistently changing.

A limitation in regards to the team fixed effects model is that the number of games a team had from 2004 to 2014 in a division playoff race is a relatively small sample size. This makes it difficult to make broad position statements. However, if the data were inclusive of a larger date range, it would be difficult to find consistency in the managers, players, and front office personnel in place in order to make any inferences. The model is effective for showing that certain teams are not responsible for scoreboard watching, and that it is not an overarching phenomenon.

Another possible limitation is that the results look at outcomes of entire teams, but baseball is a sport in which outcomes are strongly dependent on individual performance.

This makes it difficult to judge if the significance of scoreboard watching is caused by the entire team or just certain individuals on a team.

The final notable limitation is that matchups between the team and the division leader are not eliminated from the data. This means that if the division leader lost their previous game, it could have been to the team that is chasing them in the playoff race. If there are cases of momentum, injuries, or resting players, this could make it more likely that the division leader would lose consecutive games. This is not a major limitation as counterfactual observation is used to project outcomes as if a factor did not exist. As outcome of division leader is being observed for the previous day, this has no direct correlation on another team's future performance.

### **Areas for Future Research**

Since no studies of scoreboard watching have previously been undertaken at the time of this research, this study serves as a seminal work in this area of research. Due to this gap in studies, a number of suggestions for future research can be made. First, research could be expanded into the other three major professional sports leagues (National Basketball Association, National Football League, and National Hockey League). This would make it possible to determine if the time separation of a game has a strong influence in regards to scoreboard watching; while MLB plays games every day, NBA and NHL are closer to every other day and NFL is weekly. This would also allow the consideration of sports where both teams are on offense/defense at the same time (hockey and basketball) versus teams being on either offense or defense (baseball and football).

Another suggestion for future research would be to conduct qualitative studies. Interviewing players and managers within Major League Baseball (though access to such personnel may be difficult) regarding past experiences where they were unable to achieve a playoff position despite expectations of being a top tier team could provide specific reasoning as to why they falter. If these interviews brought about results that could be consistently categorized, it would be easier for sports psychologists to address the issues and make the team more competitive in future playoff races. This may be more effective if interviews were conducted right after a season in which a team missed the playoffs after being in a playoff race, whereas interviewing a few seasons after the fact or into their retirement might not lead to the same detailed responses. It would then be ideal if the same team made the playoffs in a future year to interview them again, in order to differentiate a playoff team from a non-playoff team.

Possible issues that could arise from this sort of study is that players, coaches, and front office personnel have been through a lot of media training so they might not be willing to give exact details on what led to their failures. They could fear that certain things they say in confidentiality could be released to the media and cause issues going forward with the organization. A questionnaire in which a name is not required could be useful to address this concern.

Future research could also focus on if gambling markets are properly accounting for scoreboard watching. If this is not the case, people could exploit the findings in order to gain profits. As online gambling and sports betting at places such as Las Vegas casinos are massive industries, the organizations that operate them may want to maximize the information in order to make odds that will lead to bets being placed, but leaving the

“house” with a net positive position in terms of money coming in versus money going out. In order to make well-informed decisions on gambling, all possible factors should be considered, and since literature is lacking on scoreboard watching, gamblers would be advised to factor scoreboard watching in before the odds makers do so.

Another possible research study related to scoreboard watching is to determine what else could be impacted by scoreboard watching other than winning. In this study, several control factors were used to consider scoreboard watching and its effect on winning. Possible additional factors that may be impacted by scoreboard watching are the likelihood of getting injured, fan interest both in stadiums and through media, and if players are negatively impacted by scoreboard watching/choking under pressure in free agency in terms of salary when compared to players with similar statistics.

In addition to studying other North American Professional Sports leagues, future studies could also consider the impact of scoreboard watching on individual sports. An example of this would be consideration of how a tennis player performed when they have the potential to face a top seed in the next round of the tournament. Analysis of Olympic sport performance could also be conducted. Scoreboard watching could look at performance following a World Record being set or how athletes performed in comparison to their average performance following a heat involving the top ranked athlete.

Future research could consider how the division leader performs based upon the results of the teams trailing them in the standings. This specific study does not evaluate for this because leaders know that they will qualify for the postseason if they keep winning whereas teams trailing in the division race are dependent on division leader races

to surpass them in the standings. It is also difficult to consider which team they would consider if multiple teams are in the division race. Another future study consideration would be to determine how teams trailing the Wild Card team or the Wild Card team themselves performs in relation to the previous performance of other teams.

A final suggestion for future research is to isolate teams that have been successful in prior playoff races when entering September behind the division leader. Common factors could be compared for these teams and then compared to other teams that either lost their division lead or failed to qualify for the playoffs when being within five games of the division leader at the end of the month. This could help management determine the proper methods for avoidance of choking under pressure in playoff races.

## CHAPTER VII

### CONCLUSION

Much of the prior research demonstrating distraction theory lacked a consistent research design, which the current seminal study attempted to improve upon by conducting regression analysis. In order to gain unbiased and efficient estimates, OLS was selected as the regression technique. When there are two teams competing in a matchup, team quality is a strong indicator of which team will win, so difference in team quality was necessary to include in the equation. This allowed for an indicator variable representing if the division leader won or lost to be created and indicative of if scoreboard watching is a significant phenomenon or not. In order to determine if scoreboard watching had more relevance at specific points in a season, the regression equation tested for multiple factors.

This study examines 11 years and 26,370 of Major League Baseball's game outcomes to test whether distraction, through scoreboard watching, causes teams to choke under pressure. Results indicate that scoreboard watching significantly impacts the probability of winning a game, especially in playoff races. Specifically, teams in a playoff race had a 0.158 lower probability of winning games when the division leader won its game the previous day (in comparison to a 0.097 reduced probability of winning over the entirety of a season). Consistent with distraction theory, the analysis also shows that the distraction effects are 0.224 greater on home teams. There is evidence of increased distraction as criticality of games increase. When there are fewer than 10 games remaining in a playoff race, the impact of a division leader win reduces a team's win probability by 0.243. Changes to league structure reduced win probability by 0.039 for

seasons starting in 2012. This involved the addition of a Wild Card team to each league and an increase to the value of winning a division.

This study helps fill a gap in the literature in relation to research on external factors and their impact on game outcomes. If a team has the ability to account for factors related to winning a game then it could be possible to gain a competitive advantage over the opposition. The findings also have practical applications. MLB teams can take initiatives to eliminate distraction and keep players' attention on the task at hand surrounding critical games.

Based upon these findings, it is suggested that both quantitative and qualitative studies be continued on the scoreboard watching phenomenon. For teams to improve based on the results of this study, it is suggested that they contract with sports psychologists to address scoreboard watching on a one-on-one basis with players and coaches. Athletes need to understand the information is out there but that it is outside of their control. OK to be exposed to it but over-consideration of it could lead to anxiety and suboptimal performance. Having awareness and following traditional preparation techniques could allow for success in pressure filled situations. If a team is able to eliminate this factor, it could lead to postseason qualification and additional revenue. Any postseason team has a chance to win a championship, so elimination of any factors that lead to suboptimal performance would be preferable. If teams placed more attention on their own play instead of the performance of other teams in their respective games, the possibility of additional success is more likely. A season contains 162 games, so if a team misses the playoffs by a game or two, eliminating scoreboard watching has the potential to lead to additional wins. The results of the statistical analysis undertaken in this study

exhibit scoreboard watching as a significant phenomenon. The study of this phenomenon has proven to be an important gap in current research, as scoreboard watching is a determinant of game outcomes.



## REFERENCES

- Baumeister, R. F. (1984). Choking under pressure: self-consciousness and paradoxical effects of incentives on skillful performance. *Journal of personality and social psychology, 46*(3), 610.
- Baumeister, R. F., & Showers, C. J. (1986). A review of paradoxical performance effects: Choking under pressure in sports and mental tests. *European Journal of Social Psychology, 16*(4), 361–383.
- Beilock, S. L. (2007). Understanding skilled performance: Memory, attention, and ‘choking under pressure’. In T. Morris, P. Terry & S. Gordon (Eds.), *Sport & exercise psychology: International perspectives* (pp. 153–166). Morgantown, WV: Fitness Information Technology.
- Beilock, S. L., & Carr, T. H. (2001). On the fragility of skilled performance: What governs choking under pressure?. *Journal of Experimental Psychology: General, 130*(4), 701.
- Beilock, S. L., & Gray, R. (2007). Why do athletes choke under pressure? In *Handbook of Sport Psychology* (pp. 425-444). Hoboken, NJ: Wiley.
- Beilock, S. L., Kulp, C. A., Holt, L. E., & Carr, T. H. (2004). More on the fragility of performance: choking under pressure in mathematical problem solving. *Journal of Experimental Psychology: General, 133*(4), 584.
- Berger, J., & Pope, D. (2011). Can losing lead to winning?. *Management Science, 57*(5), 817–827.
- Cao, Z., Price, J., & Stone, D. F. (2011). Performance under pressure in the NBA. *Journal of Sports Economics, 12*(3), 231–252.

- Clark, T. P., Tofler, I. R., & Lardon, M. T. (2005). The sport psychiatrist and golf. *Clinics in sports medicine*, 24(4), 959-971.
- Dohmen, T. J. (2008). Do professionals choke under pressure?. *Journal of Economic Behavior & Organization*, 65(3), 636-653.
- Fairclough, S. H., Tattersall, A. J., & Houston, K. (2006). Anxiety and performance in the British driving test. *Transportation Research Part F: Traffic Psychology and Behaviour*, 9(1), 43-52.
- Gómez, M. Á., Lorenzo, A., Jiménez, S., Navarro, R. M., & Sampaio, J. (2015). Examining choking in basketball: effects of game outcome and situational variables during last 5 minutes and overtimes. *Perceptual and motor skills*, 120(1), 111-124.
- Gray, R. (2004). Attending to the execution of a complex sensorimotor skill: Expertise differences, choking, and slumps. *Journal of Experimental Psychology: Applied*, 10(1), 42.
- Hickman, D. C., & Metz, N. E. (2015). The impact of pressure on performance: Evidence from the PGA TOUR. *Journal of Economic Behavior & Organization*, 116, 319-330.
- Hill, D. M., Hanton, S., Fleming, S., & Matthews, N. (2009). A re-examination of choking in sport. *European Journal of Sport Science*, 9(4), 203-212.
- Hill, D. M., Hanton, S., Matthews, N., & Fleming, S. (2010). A qualitative exploration of choking in elite golf. *Journal of Clinical Sport Psychology*, 4(3), 221.
- Jones, G., Hanton, S., & Connaughton, D. (2007). A framework of mental toughness in the world's best performers. *The Sport Psychologist*, 21(2), 243-264.

- Mesagno, C., Marchant, D., & Morris, T. (2009). Alleviating choking: The sounds of distraction. *Journal of Applied Sport Psychology, 21*(2), 131–147.
- Otten, M. P., & Barrett, M. E. (2013). Pitching and clutch hitting in Major League Baseball: What 109 years of statistics reveal. *Psychology of Sport and Exercise, 14*(4), 531–537.
- Pope, D., & Schweitzer, M. (2011). Is Tiger Woods loss averse? Persistent bias in the face of experience, competition, and high stakes. *American Economic Review, 101*(1), 129–157.
- Sampaio, J., Lago, C., Casais, L., & Leite, N. (2010). Effects of starting score-line, game location, and quality of opposition in basketball quarter score. *European Journal of Sport Science, 10*(6), 391–396.
- Sanders, G. S. (1981). Driven by distraction: An integrative review of social facilitation theory and research. *Journal of Experimental Social Psychology, 17*(3), 227–251.
- Sanders, S., & Walia, B. (2012). Shirking and “choking” under incentive-based pressure: a behavioral economic theory of performance production. *Economics Letters, 116*(3), 363–366.
- Schlossberg, D., Baxter, K., & Conine, J. (2004). *Miracle Over Miami: How the 2003 Marlins Shocked the World*. Miami, Florida: Sports Publishing LLC.
- Szramiak, J. (2016, April 3). Here are 5 Warren Buffett quotes about baseball and investing. *Business Insider*.
- Toma, M. (2015). Missed Shots at the Free-Throw Line Analyzing the Determinants of Choking Under Pressure. *Journal of Sports Economics, 15*27002515593779.

- Turner, M. J., & Barker, J. B. (2013). Resilience: lessons from the 2012 Olympic Games. *Reflective Practice, 14*(5), 622-631.
- Williams, A. M., & Rodrigues, J. V. S. (2002). A Test of Eysenck and Calvo's Processing Efficiency Theory. *Journal of Sport & Exercise Psychology, 24*, 438-455.
- Wine, J. (1971). Test anxiety and direction of attention. *Psychological Bulletin, 76*(2), 92.
- Worthy, D. A., Markman, A. B., & Maddox, W. T. (2009). Choking and excelling at the free throw line. *Korean Journal of Thinking and Problem Solving, 19*(1), 53.

## APPENDIX A: PYTHON CODE FOR ANALYSIS

### Import modules

In [220]:

```
import sys, os
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import pylab
import statsmodels.api as sm
import statsmodels.formula.api as smf
from statsmodels.formula.api import ols
from dateutil import parser
```

### Set display options

In [221]:

```
pd.set_option('display.max_columns', 10)
pd.set_option('display.max_rows', 10)
desired_width = 320
pd.set_option('display.width', desired_width)
```

### Import Data

In [222]:

```
currPath = '/home/vmuser/Documents/analysis'

dataManager = 'ManagerStats.csv'
dataScores = 'MLBScores.csv'
dataSalaries = 'Salaries.csv'
dataTeamStats = 'TeamStats.csv'
dataDates = 'DatesbyTeam.csv'
```

### Descriptive Statistics of Game Outcomes

In [223]:

```
dc = pd.read_csv(currPath + '/' + dataScores, low_memory=False)

dc.describe().round(2)
```

Out [223]:

	Season	Rk	Gm	Date	Unnamed: 4...	D/N	Attendance	Streak	Streak1	PriorStreak	
<b>Count</b>	54803	54803	54803	54803	53454	...	54803	54803	54803	54016	
<b>unique</b>	13	164	164	1278	1	...	3	19044	34	34	35
<b>Top</b>	2013	Rk	Gm#	Date	boxscore	...	N	Attendance	-	-1	1
<b>Freq</b>	4982	1349	1349	1349	53454	...	36064	1349	13170	13170	13304

4 rows x 26 columns

## Descriptive Statistics of Managerial Statistics

In [224]:

```
dmgr = pd.read_csv(currPath + '/' + dataManager, low_memory=False)

dmgr.describe().round(2)
```

Out [224]:

	Season	Yrs	Prior Seasons Managed	From	To	...	WSwon	PennWon	ASG	playoff per season	career playoffs per season
<b>Count</b>	330.00	330.00	330.00	330.00	330.00	...	330.00	330.00	330.00	330.00	330.00
<b>Mean</b>	2009.00	11.46	7.52	1999.51	2012.27	...	0.54	1.03	0.98	0.21	0.25
<b>Std</b>	3.17	7.48	7.23	9.67	2.41	...	0.96	1.58	1.53	0.24	0.20
<b>Min</b>	2004.00	1.00	0.00	1973.00	2004.00	...	0.00	0.00	0.00	0.00	0.00
<b>25%</b>	2006.00	6.00	2.00	1995.00	2011.00	...	0.00	0.00	0.00	0.00	0.02
<b>50%</b>	2009.00	11.00	6.00	2002.00	2014.00	...	0.00	0.00	0.00	0.17	0.25
<b>75%</b>	2012.00	15.00	11.00	2007.00	2014.00	...	1.00	1.00	1.00	0.36	0.38
<b>Max</b>	2014.00	33.00	32.00	2014.00	2014.00	...	4.00	6.00	6.00	1.00	1.00

8 rows x 23 columns

## Descriptive Statistics of Team Payroll

In [225]:

```
dpay = pd.read_csv(currPath + '/' + dataSalaries, low_memory=False)

dpay.describe().round(2)
```

Out [225]:

	Year	Start Pay	End Pay	End Rank	OWAR	DWAR	PWAR	WAR
<b>Count</b>	330.00	3.300000e+02	3.300000e+02	330.00	330.00	330.00	330.00	330.00
<b>Mean</b>	2009.00	9.009438e+07	9.511560e+07	15.48	19.68	-0.00	13.67	33.34
<b>Std</b>	3.17	3.923318e+07	4.138032e+07	8.67	6.65	3.81	7.03	10.64
<b>Min</b>	2004.00	1.499850e+07	2.112433e+07	1.00	0.50	-12.40	-5.80	3.80
<b>25%</b>	2006.00	6.329638e+07	6.920405e+07	8.00	14.82	-2.40	8.83	25.52
<b>50%</b>	2009.00	8.518380e+07	8.981920e+07	15.00	19.65	0.25	14.00	33.55
<b>75%</b>	2012.00	1.069286e+08	1.135385e+08	23.00	23.80	2.70	18.17	41.00
<b>Max</b>	2014.00	2.293359e+08	2.572834e+08	30.00	40.00	11.10	37.20	59.90

## Descriptive Statistics of Average Age

In [226]:

```
dage = pd.read_csv(currPath + '/' + dataTeamStats, low_memory=False)

dage.describe().round(2)
```

Out[226]:

	Season	NumBat	BatAge	PerBat	NumP	PAge	PerP	AvgAge
<b>Count</b>	330.00	330.00	330.00	330.00	330.00	330.00	330.00	330.00
<b>Mean</b>	2009.00	46.07	28.87	0.66	23.52	28.61	0.34	28.79
<b>Std</b>	3.17	4.65	1.33	0.02	3.49	1.38	0.02	1.20
<b>Min</b>	2004.00	36.00	25.60	0.61	15.00	25.70	0.28	25.70
<b>25%</b>	2006.00	43.00	27.90	0.65	21.00	27.70	0.32	28.03
<b>50%</b>	2009.00	46.00	28.80	0.66	23.00	28.40	0.34	28.58
<b>75%</b>	2012.00	49.00	29.70	0.68	25.75	29.37	0.35	29.43
<b>Max</b>	2014.00	64.00	33.50	0.72	40.00	34.20	0.39	33.04

## Team-game dates 2004-2014

In [227]:

```
dd = pd.read_csv(currPath + '/' + dataDates, low_memory=False)
dd['fullDate'] = dd.apply(lambda x: parser.parse(x['Date']), axis=1)
dd = dd.sort_values(by=['Season'], ascending=[1])

print("Descriptive statistics")
dd.describe().round(2)
```

Descriptive statistics

Out[227]:

	Season
<b>Count</b>	67650.00
<b>Mean</b>	2009.00
<b>Std</b>	3.16
<b>Min</b>	2004.00
<b>25%</b>	2006.00
<b>50%</b>	2009.00
<b>75%</b>	2012.00
<b>Max</b>	2014.00

## Create scores data set and import dates

In [228]:

```
dc = pd.read_csv(currPath + '/' + dataScores, low_memory=False)
dc = dc[dc['Date'] != 'Date'] #Remove extra heading rows
dc['Date'] = dc['Date'] + ', ' + dc['Season']
dc['Date'] = dc.apply(lambda x: parser.parse(x['Date']), axis=1)
```

```
dc['isthome'] = dc.apply(lambda x: 1 if x['Symbol']!="@" else 0, axis=1)
```

### Rename headers and convert values to floats

In [229]:

```
dc = dc.rename(columns={'Visitor': 'team', 'Home': 'opp', 'Visitor W/L': 'gameW
L', 'Runs Visitor': 'teamRuns', 'Runs Home': 'oppRuns', 'D/N': 'DN', 'Gm': 'gn'
})
dc = dc.drop(['Unnamed: 4', 'W-L', 'Win', 'Loss', 'Save', 'Rk', 'gameWL', 'Inn'
, 'GB', 'Rank', 'Streak', 'Time', 'Symbol'], axis=1)
dc[['Season', 'teamRuns', 'oppRuns', 'gn']] = dc[['Season', 'teamRuns', 'oppRun
s', 'gn']].astype(float)
```

### Determine winners and losers by run differential

In [264]:

```
dc['druns'] = dc['teamRuns'] - dc['oppRuns'] #difference in runs
dc['twin'] = dc.apply(lambda x: 1 if x['druns'] >0 else 0, axis=1) #winner ind
icator variable
dc['tlos'] = dc.apply(lambda x: 1 if x['druns'] < 0 else 0, axis=1) #loser indi
cator variable

dc = dc.sort_values(by=['Season', 'team', 'gn'], ascending=[1, 1, 1])
dc = dc.set_index(['Season', 'team'])
```

### Attain pre-game records

In [231]:

```
dc['tws'] = dc.groupby(level=['Season', 'team'])['twin'].cumsum() - dc['twin']
dc['tls'] = dc.groupby(level=['Season', 'team'])['tlos'].cumsum() - dc['tlos']
dc['twper'] = dc['tws'] / (dc['tws'] + dc['tls'])
dc['tgameday'] = 1
dc = dc.reset_index()
```

### Loop to create individual data by season/team

In [232]:

```
season_list = dc['Season'].unique() ###Create list of seasons
team_list = dc['team'].unique() ###Create list of teams
dm = pd.DataFrame() #####Create blank data frame

for season in season_list:
    dt = dc[dc['Season'] == season] #####Create data for one season
```



```

sta_date = dt['Date'].min()
end_date = dt['Date'].max()
for team in team_list:
    du = pd.DataFrame(data=None, columns=['team'], index=pd.date_range(sta_
date, end_date)).reset_index()
    du = du.rename(columns={'index': 'Date'})
    du['team'] = team
    dv = pd.DataFrame()
    dv = pd.merge(du, dt, left_on=['Date', 'team'], right_on=['Date', 'team
'], how='left').set_index(['Season', 'team']) ###Data for one team for one seas
on
    dm = pd.concat([dm, dv], axis=0)
print 'loop complete'

```

loop complete

### Insert team record for days with no games

In [233]:

```

dm = dm.reset_index(level=0)
dm['toffday'] = dm['tgameday'].apply(lambda x: 1 if x != 1 else 0)
dm = dm.fillna(method='bfill')

```

### Include game outcomes only from May 1 to the end of the season

In [234]:

```

dm['month'] = dm['Date'].apply(lambda x: x.month)
dm = dm[dm['month'] > 4] ###Only leave May results to the end of the season
results
dm = dm.reset_index()

```

### Create League Standings

In [263]:

```

dr = pd.DataFrame()
dr = dm[['Season', 'Date', 'League', 'Division', 'team', 'twin', 'tlos', 'gn',
'tws', 'tls', 'twper', 'Attendance', 'Streak1', 'PriorStreak']]
#dr = dr[dr['Season']==2014]
#dr = dr[(dr['League']=='AL') & (dr['Division']=='East')]
dr = dr.sort_values(by=['League', 'Division', 'Date', 'team', 'gn'], ascending=
[1, 1, 1, 0, 1])
dr = dr.groupby(['League', 'Division', 'Date', 'team']).first().reset_index()

```

## Create League and Division Rank Variables

In [262]:

```
dr = dr.sort_values(by=['League', 'Division', 'Date', 'twper'], ascending=[1, 1, 1, 0])

dr['drank'] = dr.groupby(['League', 'Division', 'Date'])['twper'].rank(ascending = False)

dr = dr.sort_values(by=['League', 'Date', 'twper'], ascending=[1, 1, 0])

dr['lrank'] = dr.groupby(['League', 'Date'])['twper'].rank(ascending = False)
```

## Create data frame for Team and Opponent

In [237]:

```
dt = dr[['Season', 'Date', 'League', 'Division', 'team', 'drank', 'lrank']]

dt = dt.rename(columns={'drank': 'tdrank', 'lrank': 'tlrank'})

do = dr

do = do.rename(columns={'team': 'opp', 'tws': 'ows', 'tls': 'ols', 'twper': 'owper', 'drank': 'odrank', 'lrank': 'olrank', 'gn': 'ogn', 'Streak1': 'ostreak', 'PriorStreak': 'opriorstreak'})
```

## Create data frame for Division and League Leader

In [238]:

```
dw = dr[dr['drank']==1][['Season', 'Date', 'League', 'Division', 'team', 'tws', 'tls', 'twper', 'twin', 'tlos']] ###division leader data

dw = dw.rename(columns={'team': 'dleader', 'tws': 'dlws', 'tls': 'dlls', 'twper': 'dlwper', 'twin': 'dltwin', 'tlos': 'dllos'})

dx = dr[dr['lrank']==1][['Season', 'Date', 'League', 'Division', 'team', 'tws', 'tls', 'twper', 'twin', 'tlos']] ####league leader data

dx = dx.rename(columns={'team': 'lleader', 'tws': 'llws', 'tls': 'llls', 'twper': 'llwper', 'twin': 'lltwin', 'tlos': 'lllos'})
```

## Create data set for analysis

In [239]:

```
da = pd.DataFrame()

da = dm[dm['tgameday']==1][['Season', 'Date', 'League', 'Division', 'team', 'opp', 'teamRuns', 'oppRuns', 'isthome', 'druns', 'twin', 'tws', 'tls', 'twper', 'DN', 'PriorStreak']]
```

## Merge Team, Opponent, Division leader, and League Leader data sets

In [240]:

```
da = pd.merge(da, dt, left_on=['Season', 'League', 'Division', 'Date', 'team'], right_on=['Season', 'League', 'Division', 'Date', 'team'])
```

```

da = pd.merge(da, do, left_on=['Season', 'Date', 'opp'] , right_on=['Season', 'Date', 'opp'])

da = pd.merge(da, dw, left_on=['Season', 'League_x', 'Division_x', 'Date'], right_on=['Season', 'League', 'Division', 'Date'])

da = pd.merge(da, dx, left_on=['Season', 'League_x', 'Date'], right_on=['Season', 'League', 'Date'])

```

### Create games back variables

In [241]:

```

'''playoff race indicative variables'''
da['gbl1'] = (da['llws']-da['tws'] + da['tls']-da['lls'])/2
da['gbd1'] = (da['dlws']-da['tws'] + da['tls']-da['dls'])/2

da = da.rename(columns={'League_x': 'tLeague', 'Division_x': 'tDivision', 'League_y': 'oLeague', 'Division_y': 'oDivision'})

```

### Attain previous day's game outcome for division leader

In [242]:

```

import datetime
from datetime import timedelta
dw['one_day'] = datetime.timedelta(days=1)
dw['nextday'] = dw['Date']+dw['one_day']
dw['Date'].dtypes

```

Out[242]:

### Merge previous division leader game outcome with team result

In [243]:

```

d1 = pd.merge(da, dw, left_on=['Season', 'Date', 'dleader'], right_on=['Season', 'nextday', 'dleader'])
d1.dtypes
d1.describe()
d1.groupby(level=0).first()
print 'merge complete'

```

merge complete

### Merge October results with September

In [244]:

```

d1['month'] = d1['Date_x'].apply(lambda x: x.month)

```

```
d1['month2'] = d1['month']
d1['month2'] = d1.apply(lambda x: 9 if (x['month2'] > 9) else x['month2'] , axis=1)
```

### Create variables representing games back by category

In [245]:

```
d1['gb'] = d1.apply(lambda x: 2 if (x['gbd1'] > 10) else x['gbd1'] , axis=1) #more than 10 games back
d1['gb'] = d1.apply(lambda x: 0 if (x['gbd1'] <= 5) else x['gb'] , axis=1) #5 to 10 games back
d1['gb'] = d1.apply(lambda x: 1 if (x['gbd1'] > 5) & (x['gbd1'] <= 10) else x['gb'] , axis=1) #less than 5 games back
```

### Create data set that excludes division leader game outcomes

In [246]:

```
d1 = d1[d1['gbd1'] > 0]
```

### Convert variable types to floats

In [247]:

```
d1['Season'] = d1['Season'].astype(float)
d1['month'] = d1['month'].astype(float)
d1['gb'] = d1['gb'].astype(float)
d1['dltwin_x'] = d1['dltwin_x'].astype(float)
d1['twin_x'] = d1['twin_x'].astype(float)
d1['month2'] = d1['month2'].astype(float)
d1['PriorStreak'] = d1['PriorStreak'].astype(float)
d1['opriorstreak'] = d1['opriorstreak'].astype(float)
```

### Table 1: Summary Statistics of Scoreboard Watching

In [248]:

```
d3 = d1.groupby(['month2', 'gb', 'dltwin_x'])['twin_x'].mean()
print d3
d4 = d1.groupby(['month2', 'gb', 'dltwin_x'])['twin_x'].std()
print d4
print 'gb = 0: less than 5 games back'
print 'gb = 1: 5 to 10 games back'
print 'gb = 2: greater than 10 games back'
```

```
month2 gb dltwin_x
```

```

5.0      0.0  0.0      0.566518
          1.0      0.473811
          1.0  0.0      0.494764
          1.0      0.451777
          2.0  0.0      0.533724
          ...
9.0      0.0  1.0      0.506045
          1.0  0.0      0.625187
          1.0      0.454128
          2.0  0.0      0.542974
          1.0      0.380597
Name: twin_x, dtype: float64
month2  gb    dltwin_x
5.0      0.0  0.0      0.495713
          1.0      0.499464
          1.0  0.0      0.500191
          1.0      0.497880
          2.0  0.0      0.499594
          ...
9.0      0.0  1.0      0.500396
          1.0  0.0      0.484438
          1.0      0.498177
          2.0  0.0      0.498263
          1.0      0.485624
Name: twin_x, dtype: float64
gb = 0: less than 5 games back
gb = 1: 5 to 10 games back
gb = 2: greater than 10 games back

```

### Create additional variables for analysis

In [249]:

```

d1['lngbdl'] = np.log(d1['gbdl']) #log of games back
d1['dwper'] = d1['twper'] - d1['owper'] #difference in winning percentage w
ith opponent
d1['dlw_gbdl'] = d1['dltwin_x'] * d1['gbdl'] #games back of division leader if
division leader won
d1['dll_gbdl'] = (1 - d1['dltwin_x']) * d1['gbdl'] #games back of division lea
der if division leader lost

d1['twperhalf'] = d1['twper']-0.5 #team winning percentage minus 50%

Y = d1['twin_x'] #Create dependent variable

```

### Table 2: Ordinary Least Squares in relation to Scoreboard Watching and Win Probability

In [250]:

```

print "all seasons"
print 'all months'
X = sm.add_constant(d1[['dltwin_x', 'dwper']])
tempOut = sm.OLS(Y, X).fit()
print tempOut.summary()

```

```

print 'May to August'
dt = d1[d1['month2']<9]
Y = dt['twin_x']
X = sm.add_constant(dt[['dltwin_x', 'dwper']])
tempOut = sm.OLS(Y, X).fit()
print tempOut.summary()

print 'September to October'
ds = d1[d1['month2']>=9]
Y = ds['twin_x']
X = sm.add_constant(ds[['dltwin_x', 'dwper']])
tempOut = sm.OLS(Y, X).fit()
print tempOut.summary()

print 'Teams still in race during September/October'
ds['gr'] = 162-ds['tws']-ds['tls']
dsi = ds[ds['gr']>=ds['gbd1']]

Y = dsi['twin_x']
X = sm.add_constant(dsi[['dltwin_x', 'dwper']])
tempOut = sm.OLS(Y, X).fit()
print tempOut.summary()

print 'Teams still in race during September/October based on games back'
Y = dsi['twin_x']
X = sm.add_constant(dsi[['dltwin_x', 'dwper', 'dlw_gbd1', 'dll_gbd1']])
tempOut = sm.OLS(Y, X).fit()
print tempOut.summary()

```

```

all seasons
all months

```

#### OLS Regression Results

```

=====
Dep. Variable:          twin_x    R-squared:                0.020
Model:                  OLS       Adj. R-squared:           0.020
Method:                 Least Squares   F-statistic:              368.8
Date:                   Thu, 08 Dec 2016   Prob (F-statistic):       3.07e-159
Time:                   18:21:19         Log-Likelihood:           -25309.
No. Observations:      35386           AIC:                      5.062e+04
Df Residuals:          35383           BIC:                      5.065e+04
Df Model:               2
Covariance Type:       nonrobust
=====

```

	coef	std err	t	P> t	[95.0% Conf. Int.]	
const	0.5547	0.004	140.690	0.000	0.547	0.562
dltwin_x	-0.0972	0.005	-18.413	0.000	-0.108	-0.087
dwper	0.5169	0.026	19.910	0.000	0.466	0.568

=====

Omnibus:	8.276	Durbin-Watson:	2.174
Prob(Omnibus):	0.016	Jarque-Bera (JB):	5427.486
Skew:	0.037	Prob(JB):	0.00
Kurtosis:	1.083	Cond. No.	11.5

=====

## Warnings:

[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.

May to August

## OLS Regression Results

=====

Dep. Variable:	twin_x	R-squared:	0.016
Model:	OLS	Adj. R-squared:	0.016
Method:	Least Squares	F-statistic:	224.9
Date:	Thu, 08 Dec 2016	Prob (F-statistic):	1.30e-97
Time:	18:21:19	Log-Likelihood:	-20032.
No. Observations:	27915	AIC:	4.007e+04
Df Residuals:	27912	BIC:	4.009e+04
Df Model:	2		
Covariance Type:	nonrobust		

=====

	coef	std err	t	P> t	[95.0% Conf. Int.]	
const	0.5453	0.004	122.964	0.000	0.537	0.554
dltwin_x	-0.0809	0.006	-13.576	0.000	-0.093	-0.069
dwper	0.4680	0.029	16.283	0.000	0.412	0.524

=====

Omnibus:	5.805	Durbin-Watson:	2.146
Prob(Omnibus):	0.055	Jarque-Bera (JB):	4363.600
Skew:	0.035	Prob(JB):	0.00
Kurtosis:	1.064	Cond. No.	11.3

=====

## Warnings:

[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.

September to October

## OLS Regression Results

=====

Dep. Variable:	twin_x	R-squared:	0.044
Model:	OLS	Adj. R-squared:	0.044
Method:	Least Squares	F-statistic:	172.7
Date:	Thu, 08 Dec 2016	Prob (F-statistic):	4.79e-74
Time:	18:21:19	Log-Likelihood:	-5250.2
No. Observations:	7471	AIC:	1.051e+04
Df Residuals:	7468	BIC:	1.053e+04
Df Model:	2		
Covariance Type:	nonrobust		

=====

	coef	std err	t	P> t	[95.0% Conf. Int.]	
const	0.5912	0.009	68.887	0.000	0.574	0.608
dltwin_x	-0.1581	0.011	-13.898	0.000	-0.180	-0.136

```

dwper          0.7358      0.060      12.178      0.000      0.617      0.854
=====
Omnibus:                1.814      Durbin-Watson:                2.293
Prob(Omnibus):          0.404      Jarque-Bera (JB):            1033.957
Skew:                   0.038      Prob(JB):                    3.01e-225
Kurtosis:               1.179      Cond. No.                     12.5
=====

```

## Warnings:

[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.

Teams still in race during September/October

## OLS Regression Results

```

=====
Dep. Variable:          twin_x      R-squared:                0.042
Model:                 OLS         Adj. R-squared:           0.042
Method:                Least Squares  F-statistic:              84.38
Date:                  Thu, 08 Dec 2016  Prob (F-statistic):      1.37e-36
Time:                  18:21:19      Log-Likelihood:          -2703.2
No. Observations:     3839         AIC:                     5412.
Df Residuals:         3836         BIC:                     5431.
Df Model:              2
Covariance Type:      nonrobust
=====

```

```

-----
                coef      std err          t      P>|t|      [95.0% Conf. Int.]
-----
const           0.5836      0.012      48.675      0.000      0.560      0.607
dltwin_x       -0.1470      0.016     -9.228      0.000     -0.178     -0.116
dwper           0.8741      0.097      8.982      0.000      0.683      1.065
=====

```

```

Omnibus:                1.642      Durbin-Watson:                2.258
Prob(Omnibus):          0.440      Jarque-Bera (JB):            536.748
Skew:                   -0.051      Prob(JB):                    2.80e-117
Kurtosis:               1.171      Cond. No.                     14.5
=====

```

## Warnings:

[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.

Teams still in race during September/October based on games back

## OLS Regression Results

```

=====
Dep. Variable:          twin_x      R-squared:                0.043
Model:                 OLS         Adj. R-squared:           0.042
Method:                Least Squares  F-statistic:              43.25
Date:                  Thu, 08 Dec 2016  Prob (F-statistic):      1.53e-35
Time:                  18:21:19      Log-Likelihood:          -2701.1
No. Observations:     3839         AIC:                     5412.
Df Residuals:         3834         BIC:                     5444.
Df Model:              4
Covariance Type:      nonrobust
=====

```

```

-----
                coef      std err          t      P>|t|      [95.0% Conf. Int.]
-----
const           0.6178      0.024     26.115      0.000      0.571      0.664
dltwin_x       -0.1548      0.030     -5.153      0.000     -0.214     -0.096
=====

```



```

dwper          0.7674      0.111      6.896      0.000      0.549      0.986
dlw_gbd1      -0.0026      0.002      -1.384     0.166     -0.006     0.001
dll_gbd1      -0.0033      0.002      -1.653     0.098     -0.007     0.001
=====
Omnibus:                1.486      Durbin-Watson:                2.257
Prob(Omnibus):          0.476      Jarque-Bera (JB):             534.399
Skew:                   -0.048     Prob(JB):                     9.05e-117
Kurtosis:               1.175     Cond. No.                     125.
=====

```

## Warnings:

```
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
```

Table 3: Scoreboard Watching Regression Based on Playoff Races

In [251]:

```

print 'August in race'
da = d1[d1['month2']==8]
da = da[da['gbd1']<=5]
Y = da['twin_x']
X = sm.add_constant(da[['dltwin_x', 'dwper']])
tempOut = sm.OLS(Y, X).fit()
print tempOut.summary()

print 'September in race'
dsi5 = dsi[dsi['gbd1']<=5]
Y = dsi5['twin_x']
X = sm.add_constant(dsi5[['dltwin_x', 'dwper']])
tempOut = sm.OLS(Y, X).fit()
print tempOut.summary()

```

## August in race

## OLS Regression Results

```

=====
Dep. Variable:          twin_x      R-squared:                0.015
Model:                  OLS        Adj. R-squared:           0.014
Method:                 Least Squares  F-statistic:              12.18
Date:                   Thu, 08 Dec 2016  Prob (F-statistic):       5.65e-06
Time:                   18:21:20      Log-Likelihood:          -1141.5
No. Observations:      1590         AIC:                     2289.
Df Residuals:          1587         BIC:                     2305.
Df Model:               2
Covariance Type:       nonrobust
=====

```

	coef	std err	t	P> t	[95.0% Conf. Int.]	
const	0.5473	0.021	26.095	0.000	0.506	0.588
dltwin_x	-0.0962	0.025	-3.789	0.000	-0.146	-0.046
dwper	0.4991	0.164	3.044	0.002	0.178	0.821

```

=====
Omnibus:                0.658      Durbin-Watson:                2.108
Prob(Omnibus):          0.720      Jarque-Bera (JB):             248.989

```

```
Skew: -0.050 Prob(JB): 8.57e-55
Kurtosis: 1.064 Cond. No. 15.8
=====
```

## Warnings:

[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.

September in race

## OLS Regression Results

```
=====
Dep. Variable: twin_x R-squared: 0.019
Model: OLS Adj. R-squared: 0.017
Method: Least Squares F-statistic: 9.721
Date: Thu, 08 Dec 2016 Prob (F-statistic): 6.59e-05
Time: 18:21:20 Log-Likelihood: -704.59
No. Observations: 995 AIC: 1415.
Df Residuals: 992 BIC: 1430.
Df Model: 2
Covariance Type: nonrobust
=====
```

	coef	std err	t	P> t	[95.0% Conf. Int.]	
const	0.6004	0.026	22.972	0.000	0.549	0.652
dltwin_x	-0.1210	0.031	-3.853	0.000	-0.183	-0.059
dwper	0.5255	0.227	2.311	0.021	0.079	0.972

```
=====
Omnibus: 10.541 Durbin-Watson: 2.064
Prob(Omnibus): 0.005 Jarque-Bera (JB): 153.660
Skew: -0.254 Prob(JB): 4.30e-34
Kurtosis: 1.143 Cond. No. 17.1
=====
```

## Warnings:

[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.

**Table 4: Scoreboard Watching Regression with Team Fixed Effects**

In [252]:

```
print 'September in race with team fixed effects'
namesList = pd.get_dummies(dsi5['team'], prefix='TX') #indicator variables
dsi5 = dsi5.join(namesList) #merge with dsi5

Y = dsi5['twin_x']

X = sm.add_constant(dsi5[['dltwin_x', 'dwper', 'TX_ATL', 'TX_BAL', 'TX_BOS', 'TX_CIN', 'TX_CLE', 'TX_COL', 'TX_CHW', 'TX_DET', 'TX_HOU', 'TX_KCR', 'TX_LAA', 'TX_LAD', 'TX_MIA', 'TX_MIL', 'TX_MIN', 'TX_NYM', 'TX_NYY', 'TX_OAK', 'TX_PHI', 'TX_PIT', 'TX_SDP', 'TX_SFG', 'TX_STL', 'TX_TBR', 'TX_TEX', 'TX_WSN']])

tempOut = sm.OLS(Y, X).fit()
print tempOut.summary()
```

September in race with team fixed effects

## OLS Regression Results

```
=====
Dep. Variable: twin_x R-squared: 0.043
Model: OLS Adj. R-squared: 0.015
Method: Least Squares F-statistic: 1.539
=====
```

Date: Thu, 08 Dec 2016 Prob (F-statistic): 0.0371  
Time: 18:21:20 Log-Likelihood: -692.54  
No. Observations: 995 AIC: 1443.  
Df Residuals: 966 BIC: 1585.  
Df Model: 28  
Covariance Type: nonrobust

```
=====
              coef      std err          t      P>|t|      [95.0% Conf. Int.]
-----+-----+-----+-----+-----+-----
const                0.6356      0.110      5.794      0.000      0.420      0.851
dltwin_x            -0.1165      0.032     -3.659      0.000     -0.179     -0.054
dwper                0.4917      0.239      2.059      0.040      0.023      0.960
TX_ATL              -0.1151      0.140     -0.824      0.410     -0.389      0.159
TX_BAL               0.1219      0.157      0.778      0.437     -0.186      0.430
TX_BOS              -0.0635      0.122     -0.522      0.602     -0.302      0.175
TX_CIN              -0.0310      0.143     -0.216      0.829     -0.312      0.250
TX_CLE              -0.0549      0.140     -0.393      0.695     -0.329      0.219
TX_COL               0.0331      0.128      0.259      0.796     -0.218      0.284
TX_CHW              -0.1757      0.138     -1.278      0.202     -0.446      0.094
TX_DET              -0.0207      0.138     -0.151      0.880     -0.291      0.249
TX_HOU               0.0686      0.196      0.350      0.727     -0.317      0.454
TX_KCR              -0.0578      0.167     -0.347      0.729     -0.385      0.269
TX_LAA              -0.0543      0.126     -0.431      0.666     -0.301      0.193
TX_LAD              -0.0112      0.135     -0.083      0.934     -0.275      0.253
TX_MIA              -0.3817      0.245     -1.555      0.120     -0.863      0.100
TX_MIL              -0.1582      0.132     -1.196      0.232     -0.418      0.101
TX_MIN               0.0435      0.123      0.354      0.723     -0.197      0.284
TX_NYM               0.1101      0.189      0.582      0.561     -0.261      0.482
TX_NYY               0.0283      0.131      0.217      0.829     -0.228      0.285
TX_OAK              -0.0484      0.126     -0.385      0.700     -0.295      0.199
TX_PHI              -0.0619      0.125     -0.493      0.622     -0.308      0.184
TX_PIT               0.1414      0.135      1.048      0.295     -0.123      0.406
TX_SDP              -0.0141      0.126     -0.111      0.912     -0.262      0.234
TX_SFG              -0.0606      0.120     -0.507      0.612     -0.295      0.174
TX_STL              -0.1579      0.184     -0.860      0.390     -0.518      0.202
TX_TBR              -0.0569      0.143     -0.398      0.691     -0.338      0.224
TX_TEX              -0.1226      0.133     -0.919      0.358     -0.384      0.139
TX_WSN               0.3143      0.505      0.622      0.534     -0.677      1.305
=====
Omnibus:                10.034      Durbin-Watson:                2.055
Prob(Omnibus):           0.007      Jarque-Bera (JB):              140.058
Skew:                   -0.248      Prob(JB):                      3.86e-31
Kurtosis:                1.230      Cond. No.                      45.2
=====
```

Warnings:

[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.

Table 5: Scoreboard Watching Regression Based on Home/Away

In [253]:

```
print 'September in race before 2012'
db2012 = dsi5[dsi5['Season']<2012]
Y = db2012['twin_x']
X = sm.add_constant(db2012[['dltwin_x', 'dwper']])
tempOut = sm.OLS(Y, X).fit()
print tempOut.summary()
```

```

print 'September in race 2012 to 2014'
da2012 = dsi5[dsi5['Season']>=2012]
Y = da2012['twin_x']
X = sm.add_constant(da2012[['dltwin_x', 'dwper']])
tempOut = sm.OLS(Y, X).fit()
print tempOut.summary()

```

September in race before 2012

OLS Regression Results

```

=====
Dep. Variable:          twin_x    R-squared:                0.014
Model:                  OLS      Adj. R-squared:           0.012
Method:                 Least Squares  F-statistic:              5.407
Date:                   Thu, 08 Dec 2016  Prob (F-statistic):      0.00466
Time:                   18:21:20    Log-Likelihood:          -527.04
No. Observations:      741        AIC:                     1060.
Df Residuals:          738        BIC:                     1074.
Df Model:               2
Covariance Type:      nonrobust
=====

```

	coef	std err	t	P> t	[95.0% Conf. Int.]	
const	0.6035	0.030	19.843	0.000	0.544	0.663
dltwin_x	-0.1126	0.036	-3.087	0.002	-0.184	-0.041
dwper	0.3159	0.255	1.238	0.216	-0.185	0.817

```

=====
Omnibus:                7.091    Durbin-Watson:           2.028
Prob(Omnibus):          0.029    Jarque-Bera (JB):       116.626
Skew:                   -0.241    Prob(JB):                4.73e-26
Kurtosis:               1.117    Cond. No.                16.5
=====

```

Warnings:

[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.

September in race 2012 to 2014

OLS Regression Results

```

=====
Dep. Variable:          twin_x    R-squared:                0.049
Model:                  OLS      Adj. R-squared:           0.041
Method:                 Least Squares  F-statistic:              6.445
Date:                   Thu, 08 Dec 2016  Prob (F-statistic):      0.00186
Time:                   18:21:20    Log-Likelihood:          -175.41
No. Observations:      254        AIC:                     356.8
Df Residuals:          251        BIC:                     367.4
Df Model:               2
Covariance Type:      nonrobust
=====

```

	coef	std err	t	P> t	[95.0% Conf. Int.]	
const	0.5873	0.051	11.509	0.000	0.487	0.688
dltwin_x	-0.1522	0.062	-2.469	0.014	-0.274	-0.031
dwper	1.4156	0.506	2.798	0.006	0.419	2.412

```

=====
Omnibus:                3.882    Durbin-Watson:           2.146

```

```

Prob(Omnibus):          0.144    Jarque-Bera (JB):          34.889
Skew:                  -0.301    Prob(JB):                  2.65e-08
Kurtosis:              1.287    Cond. No.                  19.6
=====

```

## Warnings:

```
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
```

Table 6: Scoreboard Watching Regression Based on Playoff Structure

In [254]:

```

print 'September in race: home team'

dhome = dsi5[dsi5['isthome']==1]
Y = dhome['twin_x']
X = sm.add_constant(dhome[['dltwin_x', 'dwper']])
tempOut = sm.OLS(Y, X).fit()
print tempOut.summary()

print 'September in race: away team'
daway = dsi5[dsi5['isthome']==0]
Y = daway['twin_x']
X = sm.add_constant(daway[['dltwin_x', 'dwper']])
tempOut = sm.OLS(Y, X).fit()
print tempOut.summary()

```

September in race: home team

## OLS Regression Results

```

=====
Dep. Variable:          twin_x    R-squared:                0.059
Model:                  OLS      Adj. R-squared:           0.055
Method:                 Least Squares
Date:                   Thu, 08 Dec 2016
Time:                   18:21:20
No. Observations:      487      Prob (F-statistic):       4.49e-07
Df Residuals:          484      Log-Likelihood:           -333.52
Df Model:               2        AIC:                      673.0
Covariance Type:       nonrobust BIC:                      685.6
=====

```

```

=====
              coef    std err          t      P>|t|      [95.0% Conf. Int.]
-----
const          0.6670     0.036    18.604     0.000         0.597     0.737
dltwin_x      -0.2326     0.044    -5.331     0.000        -0.318    -0.147
dwper          0.4094     0.321     1.277     0.202        -0.220     1.039
=====

```

```

Omnibus:                6.346    Durbin-Watson:            2.027
Prob(Omnibus):          0.042    Jarque-Bera (JB):         62.949
Skew:                   -0.281    Prob(JB):                 2.14e-14
Kurtosis:               1.330    Cond. No.                  16.9
=====

```

## Warnings:

[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.

September in race: away team

#### OLS Regression Results

Dep. Variable:	twin_x	R-squared:	0.006		
Model:	OLS	Adj. R-squared:	0.002		
Method:	Least Squares	F-statistic:	1.492		
Date:	Thu, 08 Dec 2016	Prob (F-statistic):	0.226		
Time:	18:21:20	Log-Likelihood:	-364.32		
No. Observations:	508	AIC:	734.6		
Df Residuals:	505	BIC:	747.3		
Df Model:	2				
Covariance Type:	nonrobust				
=====					
	coef	std err	t	P> t	[95.0% Conf. Int.]
-----					
const	0.5280	0.038	13.925	0.000	0.453 0.602
dltwin_x	-0.0086	0.045	-0.190	0.849	-0.097 0.080
dwper	0.5530	0.320	1.727	0.085	-0.076 1.182
=====					
Omnibus:	3.922	Durbin-Watson:	1.934		
Prob(Omnibus):	0.141	Jarque-Bera (JB):	82.737		
Skew:	-0.214	Prob(JB):	1.08e-18		
Kurtosis:	1.070	Cond. No.	17.4		
=====					

Warnings:

[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.

**Table 7: Scoreboard Watching Regression Based on Games Remaining**

In [255]:

```
print 'September in race with more than 20 games remaining'
d20 = dsi5[dsi5['gr']>20]
Y = d20['twin_x']
X = sm.add_constant(d20[['dltwin_x', 'dwper']])
tempOut = sm.OLS(Y, X).fit()
print tempOut.summary()

print 'September in race with more than 10 and less than 20 games remaining'
d10 = dsi5[dsi5['gr']>10]
d10 = d10[d10['gr']<=20]
Y = d10['twin_x']
X = sm.add_constant(d10[['dltwin_x', 'dwper']])
tempOut = sm.OLS(Y, X).fit()
print tempOut.summary()

print 'September in race with 10 or fewer games remaining'
```

```
dend = dsi5[dsi5['gr']<=10]
Y = dend['twin_x']
X = sm.add_constant(dend[['dltwin_x', 'dwper']])
tempOut = sm.OLS(Y, X).fit()
print tempOut.summary()
```

September in race with more than 20 games remaining

OLS Regression Results

```
=====
Dep. Variable:          twin_x    R-squared:                0.007
Model:                  OLS      Adj. R-squared:           0.001
Method:                 Least Squares  F-statistic:              1.128
Date:                   Thu, 08 Dec 2016  Prob (F-statistic):       0.325
Time:                   18:21:20     Log-Likelihood:           -215.40
No. Observations:      303          AIC:                     436.8
Df Residuals:          300          BIC:                     447.9
Df Model:               2
Covariance Type:       nonrobust
=====
```

	coef	std err	t	P> t	[95.0% Conf. Int.]	
const	0.6305	0.048	13.226	0.000	0.537	0.724
dltwin_x	-0.0532	0.058	-0.924	0.356	-0.166	0.060
dwper	-0.4375	0.405	-1.079	0.281	-1.235	0.360

```
=====
Omnibus:                4.473    Durbin-Watson:            1.962
Prob(Omnibus):          0.107    Jarque-Bera (JB):         49.155
Skew:                   -0.297   Prob(JB):                  2.12e-11
Kurtosis:               1.118    Cond. No.                  16.8
=====
```

Warnings:

[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.

September in race with more than 10 and less than 20 games remaining

OLS Regression Results

```
=====
Dep. Variable:          twin_x    R-squared:                0.021
Model:                  OLS      Adj. R-squared:           0.016
Method:                 Least Squares  F-statistic:              3.992
Date:                   Thu, 08 Dec 2016  Prob (F-statistic):       0.0192
Time:                   18:21:20     Log-Likelihood:           -269.28
No. Observations:      379          AIC:                     544.6
Df Residuals:          376          BIC:                     556.4
Df Model:               2
Covariance Type:       nonrobust
=====
```

	coef	std err	t	P> t	[95.0% Conf. Int.]	
const	0.5368	0.043	12.346	0.000	0.451	0.622
dltwin_x	-0.0681	0.051	-1.333	0.183	-0.168	0.032
dwper	0.9091	0.377	2.412	0.016	0.168	1.650

```
=====
Omnibus:                2.614    Durbin-Watson:            1.998
Prob(Omnibus):          0.271    Jarque-Bera (JB):         58.134
Skew:                   -0.202   Prob(JB):                  2.38e-13
Kurtosis:               1.124    Cond. No.                  17.4
=====
```

Warnings:

[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.

September in race with 10 or fewer games remaining

#### OLS Regression Results

```

=====
Dep. Variable:          twin_x      R-squared:                0.076
Model:                  OLS         Adj. R-squared:           0.070
Method:                 Least Squares  F-statistic:              12.66
Date:                   Thu, 08 Dec 2016  Prob (F-statistic):       5.19e-06
Time:                   18:21:20      Log-Likelihood:           -211.91
No. Observations:      313          AIC:                     429.8
Df Residuals:          310          BIC:                     441.0
Df Model:               2
Covariance Type:       nonrobust
=====

```

	coef	std err	t	P> t	[95.0% Conf. Int.]	
const	0.6459	0.044	14.547	0.000	0.559	0.733
dltwin_x	-0.2429	0.055	-4.435	0.000	-0.351	-0.135
dwper	1.1297	0.401	2.818	0.005	0.341	1.919

```

=====
Omnibus:                4.074      Durbin-Watson:           2.019
Prob(Omnibus):          0.130      Jarque-Bera (JB):        37.699
Skew:                   -0.278      Prob(JB):                 6.51e-09
Kurtosis:               1.393      Cond. No.                 17.5
=====

```

Warnings:

[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.

**Table 8: Impact of Roster's Average Age on Scoreboard Watching**

In [256]:

```

print 'Regression based on Average Age'

dsage = pd.merge(dsi5,dage, left_on=['team', 'Season'], right_on=['Tm', 'Season
'])

dsavg = pd.merge(dsage,dage, left_on=['opp', 'Season'], right_on=['Tm', 'Season
'])

dsavg['diffage'] = dsavg['AvgAge_x'] - dsavg['AvgAge_y']
dsavg['intdiffage'] = dsavg['diffage']* dsavg['dltwin_x']
dsavg['diffagebat'] = dsavg['BatAge_x'] - dsavg['BatAge_y']
dsavg['intdiffagebat'] = dsavg['diffagebat']* dsavg['dltwin_x']
dsavg['diffagepit'] = dsavg['PAge_x'] - dsavg['PAge_y']
dsavg['intdiffagepit'] = dsavg['diffagepit']* dsavg['dltwin_x']

Y = dsavg['twin_x']
X = sm.add_constant(dsavg[['dltwin_x', 'dwper', 'intdiffage']])
tempOut = sm.OLS(Y, X).fit()
print tempOut.summary()

```



```

print 'Based on Position Age'
Y = dsavg['twin_x']
X = sm.add_constant(dsavg[['dltwin_x', 'dwper', 'intdiffagebat', 'intdiffagepit
']))
tempOut = sm.OLS(Y, X).fit()
print tempOut.summary()

print 'Logs of difference in age'
dsavg['lnintdiffage']= np.log(dsavg['AvgAge_x']/dsavg['AvgAge_y'])* dsavg['dltw
in_x']
Y = dsavg['twin_x']
X = sm.add_constant(dsavg[['dltwin_x', 'dwper', 'lnintdiffage']])
tempOut = sm.OLS(Y, X).fit()
print tempOut.summary()

print 'Logs of difference in age by position'
dsavg['lnintdiffagebat']= np.log(dsavg['BatAge_x']/dsavg['BatAge_y'])* dsavg['d
lwtwin_x']
dsavg['lnintdiffagepit']= np.log(dsavg['PAge_x']/dsavg['PAge_y'])* dsavg['dltwi
n_x']
Y = dsavg['twin_x']
X = sm.add_constant(dsavg[['dltwin_x', 'dwper', 'lnintdiffagebat', 'lnintdiffag
epit']])
tempOut = sm.OLS(Y, X).fit()
print tempOut.summary()

```

## Regression based on Average Age

## OLS Regression Results

```

=====
Dep. Variable:          twin_x      R-squared:                0.019
Model:                  OLS         Adj. R-squared:           0.016
Method:                 Least Squares  F-statistic:              6.522
Date:                   Thu, 08 Dec 2016  Prob (F-statistic):       0.000228
Time:                   18:21:20      Log-Likelihood:          -704.52
No. Observations:      995          AIC:                     1417.
Df Residuals:          991          BIC:                     1437.
Df Model:               3
Covariance Type:       nonrobust
=====

```

	coef	std err	t	P> t	[95.0% Conf. Int.]
const	0.6012	0.026	22.927	0.000	0.550 0.653
dltwin_x	-0.1227	0.032	-3.864	0.000	-0.185 -0.060
dwper	0.5115	0.231	2.218	0.027	0.059 0.964
intdiffage	0.0049	0.013	0.374	0.709	-0.021 0.031

```

=====
Omnibus:                10.534      Durbin-Watson:           1.894
Prob(Omnibus):          0.005      Jarque-Bera (JB):        153.562
Skew:                   -0.254      Prob(JB):                 4.51e-34

```

Kurtosis: 1.144 Cond. No. 19.6

Warnings:

[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.

Based on Position Age

OLS Regression Results

```

Dep. Variable:          twin_x      R-squared:                0.020
Model:                  OLS         Adj. R-squared:          0.016
Method:                 Least Squares  F-statistic:             5.107
Date:                   Thu, 08 Dec 2016  Prob (F-statistic):      0.000448
Time:                   18:21:20      Log-Likelihood:         -704.09
No. Observations:      995          AIC:                    1418.
Df Residuals:          990          BIC:                    1443.
Df Model:               4
Covariance Type:       nonrobust
    
```

```

==
==
coef      std err          t      P>|t|      [95.0% Conf. Int.]
-----
const      0.6024      0.026      22.943      0.000      0.551      0.6
54
dltwin_x   -0.1265      0.032      -3.950      0.000     -0.189     -0.0
64
dwper      0.4883      0.232       2.105      0.036      0.033      0.9
43
intdiffagebat -0.0058      0.013      -0.438      0.661     -0.032      0.0
20
intdiffagepit 0.0117      0.012       1.001      0.317     -0.011      0.0
35
    
```

```

Omnibus:                10.487      Durbin-Watson:           1.895
Prob(Omnibus):           0.005      Jarque-Bera (JB):       152.973
Skew:                    -0.253      Prob(JB):                6.06e-34
Kurtosis:                1.147      Cond. No.:               27.2
    
```

Warnings:

[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.

Logs of difference in age

OLS Regression Results

```

Dep. Variable:          twin_x      R-squared:                0.019
Model:                  OLS         Adj. R-squared:          0.016
Method:                 Least Squares  F-statistic:             6.508
Date:                   Thu, 08 Dec 2016  Prob (F-statistic):      0.000232
Time:                   18:21:20      Log-Likelihood:         -704.54
No. Observations:      995          AIC:                    1417.
Df Residuals:          991          BIC:                    1437.
Df Model:               3
Covariance Type:       nonrobust
    
```

```

=
coef      std err          t      P>|t|      [95.0% Conf. Int.]
-----
]
-
    
```

```

const          0.6010    0.026    22.921    0.000    0.550    0.65
2
dltwin_x      -0.1225    0.032    -3.855    0.000    -0.185   -0.06
0
dwper         0.5137    0.231    2.227    0.026    0.061    0.96
6
lnintdiffage  0.1216    0.386    0.315    0.753    -0.635    0.87
8
=====
Omnibus:                10.535    Durbin-Watson:                1.895
Prob(Omnibus):          0.005    Jarque-Bera (JB):             153.589
Skew:                   -0.254    Prob(JB):                     4.45e-34
Kurtosis:               1.144    Cond. No.                     29.2
=====

```

Warnings:  
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.  
Logs of difference in age by position

OLS Regression Results

```

=====
Dep. Variable:          twin_x    R-squared:                0.020
Model:                  OLS       Adj. R-squared:           0.016
Method:                 Least Squares    F-statistic:              5.082
Date:                   Thu, 08 Dec 2016    Prob (F-statistic):      0.000468
Time:                   18:21:20         Log-Likelihood:          -704.14
No. Observations:      995          AIC:                     1418.
Df Residuals:          990          BIC:                     1443.
Df Model:               4
Covariance Type:       nonrobust
=====

```

```

=====
              coef    std err          t      P>|t|      [95.0% Conf. Int.]
-----+-----+-----+-----+-----+-----
const          0.6022    0.026    22.936    0.000    0.551    0.654
dltwin_x      -0.1261    0.032    -3.937    0.000   -0.189   -0.063
dwper         0.4917    0.232    2.120    0.034    0.037    0.947
lnintdiffagebat -0.1716    0.384    -0.447    0.655   -0.925    0.582
lnintdiffagepit 0.3253    0.342    0.951    0.342   -0.346    0.997
=====
Omnibus:                10.492    Durbin-Watson:                1.895
Prob(Omnibus):          0.005    Jarque-Bera (JB):             153.042
Skew:                   -0.254    Prob(JB):                     5.85e-34
Kurtosis:               1.147    Cond. No.                     32.9
=====

```

Warnings:  
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.

Table 9: Impact of Team Payroll on Scoreboard Watching

In [257]:

```
dspay = pd.merge(dsi5,dpay, left_on=['team', 'Season'], right_on=['Team', 'Year'])
```

```

dsal = pd.merge(dspay,dpay, left_on=['opp', 'Season'], right_on=['Team', 'Year'
])
dsal['diffsal'] = dsal['End Pay_x'] - dsal['End Pay_y']
dsal['intdiffsal'] = dsal['diffsal']* dsal['dltwin_x']
dsal['diffrank'] = dsal['End Rank_x'] - dsal['End Rank_y']
dsal['intdiffrank'] = dsal['diffrank']* dsal['dltwin_x']
dsal['lnintdiffsal']= np.log(dsal['End Pay_x']/dsal['End Pay_y'])* dsal['dltwin_x']
dsal['lnintdiffrank']= np.log(dsal['End Rank_x']/dsal['End Rank_y'])* dsal['dltwin_x']

print 'Interaction of scoreboard watching with team payroll'
Y = dsal['twin_x']
X = sm.add_constant(dsal[['dltwin_x', 'dwper', 'lnintdiffsal']])
tempOut = sm.OLS(Y, X).fit()
print tempOut.summary()

print 'Interaction of scoreboard watching with team payroll rank'
Y = dsal['twin_x']
X = sm.add_constant(dsal[['dltwin_x', 'dwper', 'lnintdiffrank']])
tempOut = sm.OLS(Y, X).fit()
print tempOut.summary()

```

Interaction of scoreboard watching with team payroll  
OLS Regression Results

```

=====
Dep. Variable:          twin_x      R-squared:                0.019
Model:                  OLS        Adj. R-squared:          0.016
Method:                 Least Squares  F-statistic:             6.532
Date:                   Thu, 08 Dec 2016  Prob (F-statistic):      0.000224
Time:                   18:21:20      Log-Likelihood:         -704.51
No. Observations:      995          AIC:                    1417.
Df Residuals:          991          BIC:                    1437.
Df Model:               3
Covariance Type:       nonrobust
=====

```

```

=====
|
|          coef      std err          t      P>|t|      [95.0% Conf. Int.]
|-----|-----|-----|-----|-----|-----|
| const          0.6015      0.026      22.895      0.000      0.550      0.65
| 3
| dltwin_x      -0.1231      0.032      -3.868      0.000      -0.186      -0.06
| 1
| dwper          0.5054      0.233       2.172      0.030      0.049      0.96
| 2
| lnintdiffsal  0.0155      0.038       0.413      0.680      -0.058      0.08
| 9
|-----|-----|-----|-----|-----|-----|
=====

```

```

Omnibus:                10.533    Durbin-Watson:           1.894
Prob(Omnibus):          0.005    Jarque-Bera (JB):       153.548
Skew:                   -0.254    Prob(JB):                4.55e-34
Kurtosis:               1.144    Cond. No.                17.6
=====

```

## Warnings:

[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.

Interaction of scoreboard watching with team payroll rank

## OLS Regression Results

```

=====
Dep. Variable:          twin_x    R-squared:              0.020
Model:                 OLS       Adj. R-squared:         0.017
Method:                Least Squares    F-statistic:           6.670
Date:                  Thu, 08 Dec 2016    Prob (F-statistic):    0.000185
Time:                  18:21:20          Log-Likelihood:        -704.30
No. Observations:     995             AIC:                   1417.
Df Residuals:         991             BIC:                   1436.
Df Model:              3
Covariance Type:      nonrobust
=====

```

```

==
              coef    std err          t      P>|t|     [95.0% Conf. Int
.]
-----
--
const          0.6023    0.026    22.934    0.000     0.551     0.6
54
dltwin_x      -0.1243    0.032    -3.920    0.000    -0.187    -0.0
62
dwper         0.4887    0.233     2.101    0.036     0.032     0.9
45
lnintdiffrank -0.0129    0.017    -0.759    0.448    -0.046     0.0
20
=====

```

```

Omnibus:                10.540    Durbin-Watson:           1.893
Prob(Omnibus):          0.005    Jarque-Bera (JB):       153.289
Skew:                   -0.254    Prob(JB):                5.17e-34
Kurtosis:               1.146    Cond. No.                18.0
=====

```

## Warnings:

[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.

**Table 10: Impact of Managerial Experience on Scoreboard Watching**

In [258]:

```

dman = pd.merge(dsi5,dmgr, left_on=['team', 'Season'], right_on=['Team', 'Season'])

dmanager = pd.merge(dman,dmgr, left_on=['opp', 'Season'], right_on=['Team', 'Season'])

dmanager['diffplayoff'] = dmanager['Prior Playoff_x'] - dmanager['Prior Playoff_y']

dmanager['intdiffplayoff'] = dmanager['diffplayoff']* dmanager['dltwin_x']

dmanager['diffseasons'] = dmanager['Prior Seasons Managed_x'] - dmanager['Prior Seasons Managed_y']

```

```

dmanager['intdiffseasons'] = dmanager['diffseasons']* dmanager['dltwin_x']
dmanager['diffplayper'] = dmanager['playoff per season_x'] - dmanager['playoff
per season_y']
dmanager['intdiffplayper'] = dmanager['diffplayper']* dmanager['dltwin_x']

print 'Interaction of scoreboard watching with playoff seasons'
Y = dmanager['twin_x']
X = sm.add_constant(dmanager[['dltwin_x', 'dwper', 'intdiffplayoff']])
tempOut = sm.OLS(Y, X).fit()
print tempOut.summary()

print 'Interaction of scoreboard watching with seasons managed'
Y = dmanager['twin_x']
X = sm.add_constant(dmanager[['dltwin_x', 'dwper', 'intdiffseasons']])
tempOut = sm.OLS(Y, X).fit()
print tempOut.summary()

print 'Interaction of scoreboard watching with percent of seasons in playoffs'
Y = dmanager['twin_x']
X = sm.add_constant(dmanager[['dltwin_x', 'dwper', 'intdiffplayper']])
tempOut = sm.OLS(Y, X).fit()
print tempOut.summary()

```

Interaction of scoreboard watching with playoff seasons  
OLS Regression Results

```

=====
Dep. Variable:          twin_x      R-squared:                0.019
Model:                  OLS         Adj. R-squared:           0.017
Method:                 Least Squares  F-statistic:              6.571
Date:                  Thu, 08 Dec 2016  Prob (F-statistic):       0.000212
Time:                  18:21:20      Log-Likelihood:          -706.42
No. Observations:      997          AIC:                     1421.
Df Residuals:          993          BIC:                     1440.
Df Model:               3
Covariance Type:       nonrobust
=====

```

```

=====
t.]          coef      std err          t      P>|t|      [95.0% Conf. In
-----
---
const        0.5975      0.026      22.815      0.000      0.546      0.
649
dltwin_x    -0.1204      0.032     -3.816      0.000     -0.182     -0.
058
dwper        0.5004      0.232       2.159      0.031      0.046      0.
955
intdiffplayoff  0.0042      0.004       0.942      0.346     -0.005      0.
013

```

```

=====
Omnibus:                9.946    Durbin-Watson:           1.864
Prob(Omnibus):          0.007    Jarque-Bera (JB):       153.834
Skew:                   -0.246   Prob(JB):                3.94e-34
Kurtosis:               1.140    Cond. No.                53.8
=====

```

## Warnings:

[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.

Interaction of scoreboard watching with seasons managed  
 OLS Regression Results

```

=====
Dep. Variable:          twin_x    R-squared:                0.020
Model:                  OLS       Adj. R-squared:           0.017
Method:                 Least Squares   F-statistic:              6.643
Date:                   Thu, 08 Dec 2016   Prob (F-statistic):       0.000192
Time:                   18:21:20    Log-Likelihood:          -706.31
No. Observations:      997        AIC:                     1421.
Df Residuals:          993        BIC:                     1440.
Df Model:               3
Covariance Type:       nonrobust
=====

```

```

===
              coef      std err          t      P>|t|      [95.0% Conf. In
t.]
-----
const          0.5973      0.026     22.833     0.000      0.546  0.
649
dltwin_x      -0.1214      0.032     -3.839     0.000     -0.184 -0.
059
dwper          0.5057      0.230      2.194     0.028      0.053  0.
958
intdiffseasons 0.0023      0.002      1.048     0.295     -0.002  0.
007
=====

```

```

=====
Omnibus:                9.957    Durbin-Watson:           1.868
Prob(Omnibus):          0.007    Jarque-Bera (JB):       153.697
Skew:                   -0.247   Prob(JB):                4.22e-34
Kurtosis:               1.141    Cond. No.                110.
=====

```

## Warnings:

[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.

Interaction of scoreboard watching with percent of seasons in playoffs  
 OLS Regression Results

```

=====
Dep. Variable:          twin_x    R-squared:                0.019
Model:                  OLS       Adj. R-squared:           0.016
Method:                 Least Squares   F-statistic:              6.482
Date:                   Thu, 08 Dec 2016   Prob (F-statistic):       0.000241
Time:                   18:21:20    Log-Likelihood:          -706.55
No. Observations:      997        AIC:                     1421.
Df Residuals:          993        BIC:                     1441.
Df Model:               3
Covariance Type:       nonrobust
=====

```

t.]	coef	std err	t	P> t	[95.0% Conf. In	
---						
const	0.5974	0.026	22.791	0.000	0.546	0.
649						
dltwin_x	-0.1209	0.032	-3.810	0.000	-0.183	-0.
059						
dwper	0.5023	0.233	2.159	0.031	0.046	0.
959						
intdiffplayper	0.0531	0.067	0.791	0.429	-0.079	0.
185						
=====						
Omnibus:		9.895	Durbin-Watson:		1.858	
Prob(Omnibus):		0.007	Jarque-Bera (JB):		153.954	
Skew:		-0.246	Prob(JB):		3.71e-34	
Kurtosis:		1.139	Cond. No.		17.5	
=====						

## Warnings:

[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.

**Table 11: Impact of Momentum on Scoreboard Watching**

In [259]:

```

dsi5['diffstreak'] = dsi5['PriorStreak'] - dsi5['opriorstreak']
dsi5 ['intdiffstreak'] = dsi5['diffstreak']* dsi5['dltwin_x']
dwstreak = dsi5[dsi5['PriorStreak'] > 0]
dlstreak = dsi5[dsi5['PriorStreak'] < 0]

print 'Interaction of scoreboard watching with Streaks'
Y = dsi5['twin_x']
X = sm.add_constant(dsi5[['dltwin_x', 'dwper', 'intdiffstreak']])
tempOut = sm.OLS(Y, X).fit()
print tempOut.summary()

print 'Interaction of scoreboard watching with Winning Streaks'
Y = dwstreak['twin_x']
X = sm.add_constant(dwstreak[['dltwin_x', 'dwper', 'intdiffstreak']])
tempOut = sm.OLS(Y, X).fit()
print tempOut.summary()

print 'Interaction of scoreboard watching with Losing Streaks'
Y = dlstreak['twin_x']
X = sm.add_constant(dlstreak[['dltwin_x', 'dwper', 'intdiffstreak']])
tempOut = sm.OLS(Y, X).fit()

```



```
print tempOut.summary()
```

Interaction of scoreboard watching with Streaks

OLS Regression Results

```
=====
Dep. Variable:          twin_x    R-squared:                0.020
Model:                  OLS       Adj. R-squared:           0.017
Method:                 Least Squares   F-statistic:              6.713
Date:                   Thu, 08 Dec 2016   Prob (F-statistic):      0.000174
Time:                   18:21:20         Log-Likelihood:          -704.24
No. Observations:      995             AIC:                     1416.
Df Residuals:          991             BIC:                     1436.
Df Model:               3
Covariance Type:       nonrobust
=====
```

```
=====
==
              coef    std err          t      P>|t|      [95.0% Conf. Int
.]
-----
--
const          0.5993    0.026     22.892    0.000      0.548      0.6
51
dltwin_x      -0.1166    0.032     -3.662    0.000     -0.179     -0.0
54
dwper         0.5478    0.229      2.392    0.017      0.098      0.9
97
intdiffstreak -0.0040    0.005     -0.838    0.402     -0.013      0.0
05
=====
Omnibus:                10.569    Durbin-Watson:           2.062
Prob(Omnibus):          0.005    Jarque-Bera (JB):        153.234
Skew:                   -0.254    Prob(JB):                 5.32e-34
Kurtosis:                1.146    Cond. No.                 49.9
=====
```

Warnings:

[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.

Interaction of scoreboard watching with Winning Streaks

OLS Regression Results

```
=====
Dep. Variable:          twin_x    R-squared:                0.028
Model:                  OLS       Adj. R-squared:           0.023
Method:                 Least Squares   F-statistic:              5.731
Date:                   Thu, 08 Dec 2016   Prob (F-statistic):      0.000719
Time:                   18:21:20         Log-Likelihood:          -420.31
No. Observations:      602             AIC:                     848.6
Df Residuals:          598             BIC:                     866.2
Df Model:               3
Covariance Type:       nonrobust
=====
```

```
=====
==
              coef    std err          t      P>|t|      [95.0% Conf. Int
.]
-----
--
const          0.6371    0.033     19.261    0.000      0.572      0.7
02
dltwin_x      -0.1470    0.053     -2.793    0.005     -0.250     -0.0
44
dwper         0.4823    0.283      1.705    0.089     -0.073      1.0
38
```

```
intdiffstreak   -0.0024    0.009   -0.269    0.788   -0.020    0.0
15
```

```
=====
Omnibus:                10.326   Durbin-Watson:                2.030
Prob(Omnibus):          0.006   Jarque-Bera (JB):             89.905
Skew:                   -0.325   Prob(JB):                     3.00e-20
Kurtosis:               1.222   Cond. No.                     53.1
=====
```

Warnings:

[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.

Interaction of scoreboard watching with Losing Streaks

OLS Regression Results

```
=====
Dep. Variable:          twin_x   R-squared:                0.017
Model:                  OLS      Adj. R-squared:           0.009
Method:                 Least Squares   F-statistic:              2.178
Date:                   Thu, 08 Dec 2016   Prob (F-statistic):       0.0901
Time:                   18:21:20         Log-Likelihood:           -281.03
No. Observations:      393           AIC:                     570.1
Df Residuals:          389           BIC:                     586.0
Df Model:               3
Covariance Type:       nonrobust
=====
```

```
==
```

```
coef    std err      t    P>|t|    [95.0% Conf. Int
```

```
.]
```

```
-----
```

```
--
```

```
const          0.5395    0.043    12.636    0.000    0.456    0.6
```

```
23
```

```
dltwin_x      -0.1212    0.062    -1.948    0.052   -0.244    0.0
```

```
01
```

```
dwper         0.6293    0.389     1.619    0.106   -0.135    1.3
```

```
94
```

```
intdiffstreak -0.0202    0.013    -1.610    0.108   -0.045    0.0
```

```
04
```

```
=====
```

```
Omnibus:                1.401   Durbin-Watson:                2.033
```

```
Prob(Omnibus):          0.496   Jarque-Bera (JB):             61.291
```

```
Skew:                   -0.144   Prob(JB):                     4.91e-14
```

```
Kurtosis:               1.087   Cond. No.                     47.7
=====
```

Warnings:

[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.

Table 12: Impact of Opposition on Scoreboard Watching

In [260]:

```
dvslead = dsi5[dsi5['odrank']== 1]
dvslead1 = dvslead[dvslead['opp'] == dvslead['dleader']]
dvsotther = dsi5[dsi5['odrank']> 1]
dvothdivlead = dvslead[dvslead['opp'] != dvslead['dleader']]

print 'Games vs Teams Division Leader'
```

```

Y = dvslead1['twin_x']
X = sm.add_constant(dvslead1[['dltwin_x', 'dwper']])
tempOut = sm.OLS(Y, X).fit()
print tempOut.summary()

print 'Games vs Non-Division Leaders'
Y = dvsother['twin_x']
X = sm.add_constant(dvsother[['dltwin_x', 'dwper']])
tempOut = sm.OLS(Y, X).fit()
print tempOut.summary()

print 'Games vs Other Division Leaders'
Y = dvothdivlead['twin_x']
X = sm.add_constant(dvothdivlead[['dltwin_x', 'dwper', 'intdiffstreak']])
tempOut = sm.OLS(Y, X).fit()
print tempOut.summary()

```

## Games vs Teams Division Leader

## OLS Regression Results

```

=====
Dep. Variable:          twin_x    R-squared:                0.925
Model:                  OLS      Adj. R-squared:           0.924
Method:                 Least Squares  F-statistic:              924.0
Date:                  Thu, 31 Aug 2017  Prob (F-statistic):       4.60e-85
Time:                  21:45:02     Log-Likelihood:           87.193
No. Observations:      153         AIC:                      -168.4
Df Residuals:          150         BIC:                      -159.3
Df Model:               2
Covariance Type:       nonrobust
=====

```

	coef	std err	t	P> t	[95.0% Conf. Int.]	
const	0.9379	0.027	34.840	0.000	0.885	0.991
dltwin_x	-0.9674	0.023	-42.744	0.000	-1.012	-0.923
dwper	-1.3712	1.177	-1.165	0.246	-3.697	0.954

```

=====
Omnibus:                231.320    Durbin-Watson:            2.089
Prob(Omnibus):          0.000     Jarque-Bera (JB):         13216.624
Skew:                   -6.654    Prob(JB):                  0.00
Kurtosis:                46.544    Cond. No.                  118.
=====

```

## Warnings:

[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.

## Games vs Non-Division Leaders

## OLS Regression Results

```

=====
Dep. Variable:          twin_x    R-squared:                0.003
Model:                  OLS      Adj. R-squared:           0.001
Method:                 Least Squares  F-statistic:              1.322
Date:                  Thu, 31 Aug 2017  Prob (F-statistic):       0.267
=====

```

Time: 21:45:02 Log-Likelihood: -563.41  
 No. Observations: 791 AIC: 1133.  
 Df Residuals: 788 BIC: 1147.  
 Df Model: 2  
 Covariance Type: nonrobust

```
=====
              coef      std err          t      P>|t|      [95.0% Conf. Int.]
-----+-----+-----+-----+-----+-----+-----
const          0.5323      0.034      15.653      0.000          0.466      0.599
dltwin_x       0.0225      0.036       0.634      0.526         -0.047      0.092
dwper          0.4175      0.280       1.493      0.136         -0.131      0.966
=====
Omnibus:                12.247      Durbin-Watson:                2.154
Prob(Omnibus):          0.002      Jarque-Bera (JB):            130.420
Skew:                   -0.309      Prob(JB):                     4.78e-29
Kurtosis:                1.109      Cond. No.                      18.9
=====
```

## Warnings:

[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.

## Games vs Other Division Leaders

## OLS Regression Results

```
=====
Dep. Variable:          twin_x      R-squared:                0.056
Model:                  OLS        Adj. R-squared:           0.017
Method:                 Least Squares      F-statistic:              1.434
Date:                   Thu, 31 Aug 2017    Prob (F-statistic):       0.248
Time:                   21:45:02          Log-Likelihood:           -35.447
No. Observations:      51           AIC:                      76.89
Df Residuals:          48           BIC:                      82.69
Df Model:               2
Covariance Type:      nonrobust
=====
```

```
=====
              coef      std err          t      P>|t|      [95.0% Conf. Int.]
-----+-----+-----+-----+-----+-----
const          0.3275      0.111       2.955      0.005          0.105      0.550
dltwin_x       0.1710      0.140       1.218      0.229         -0.111      0.453
dwper          -2.2819      2.075      -1.100      0.277         -6.455      1.891
=====
Omnibus:                353.239      Durbin-Watson:                1.647
Prob(Omnibus):          0.000      Jarque-Bera (JB):            6.681
Skew:                   0.072      Prob(JB):                     0.0354
Kurtosis:                1.233      Cond. No.                      34.4
=====
```

## Warnings:

[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.