DESPERATION IN SPORT

by

Aruni Tennakoon

B.Sc., University of Sri Jayewardenepura, Sri Lanka, 2000

M.Sc. in Industrial Mathematics, University of Sri Jayewardenepura, Sri Lanka, 2003

A PROJECT SUBMITTED IN PARTIAL FULFILLMENT

OF THE REQUIREMENTS FOR THE DEGREE OF

MASTER OF SCIENCE

in the Department

of

Statistics and Actuarial Science

SIMON FRASER UNIVERSITY

Summer 2011

APPROVAL

Name: Aruni Tennakoon

Degree: Master of Science

Title of Project: Desperation in Sport

Examining Committee: Dr. Yi Lu
Chair

Dr. Tim Swartz
Senior Supervisor
Simon Fraser University

Dr. Derek Bingham
Supervisor
Simon Fraser University

Dr. Boxin Tang
External Examiner
Simon Fraser University

Date Approved: ______________________

Abstract

In this project, we explore the psychological effect of the state of a playoff series on a sports team. The two leagues, the National Basketball Association (NBA) and the National Hockey League (NHL) were chosen for this analysis. The team's winning probability was considered as its strength and the effect of desperation was considered as its psychological effect in a particular game. Statistical models were developed and applied to the NBA and NHL data corresponding to the 2003 through 2011 playoffs to estimate the importance of the effect of desperation. The results indicate that primarily, the NHL teams' strength is less affected by the game situation than the NBA. Secondarily, for both NBA and NHL teams, the effect of desperation is high in situations where a team has won zero games in a series. In the NBA, the effect is large when a team is close to elimination in a series. The home team advantage affects desperation in both NBA and NHL sports.

Keywords: Home court/ice advantage, Sportsbook betting, Regression analysis, Gambler

Acknowledgements

I would like to begin by thanking my senior supervisor Dr. Tim Swartz. I am sure I do not have words to express my indebtedness to Dr. Tim Swartz. But I try to appreciate his endless support. He is a great supervisor and a kind advisor not only in academic but beyond that. Thanks for everything.

I am deeply indebted to my co-supervisor Dr. Derek Bingham for his support and encouragement. I thank my examining committee member Dr. Boxin Tang for his valuable comments on the project. I also thank the chair of the examining committee Dr. Yi Lu.

I offer my sincere gratitude to Dr. Joan Hu, Dr. Jinko Graham, Dr. Tom Loughin and Ian Bercovitz for helping me in many ways during my wonderful life at SFU.

There are many others in the department who have helped me reach this goal. I would like to thank all of the graduate students in the department for their friendship. I am grateful to Sadika, Kelly and Charlene for their kindness and assistance.

I would like to express my heartfelt thanks to my son Mahen, who was very patient and helped me to work on my research, and for my parents, sister and relatives who were keeping me up time to time and make my mind fresh over the last two years.

Finally, I appreciate my husband Saman Muthukumarana for his patient support and understanding. He is a rare breed.

Contents

Approval ii

Abstract iii

Acknowledgements iv

Contents v

List of Figures vii

List of Tables viii

1 Introduction 1

1.1 What is Desperation in Sports? . . . 1

1.2 Background of the Project . . . 1

1.3 Organization of the Project . . . 2

2 Model Development 3

2.1 An Explanation of Sports Betting Lines . . . 3

2.2 Converting Betting Lines to Probabilities . . . 5

2.3 Statistical Models . . . 7

3 Data Analysis 12

3.1 NBA Data . . . 12

3.2 NHL Data . . . 15

4 Conclusions 18

4.1 Discussion 18
4.2 Future Work 19

Appendices **19**

A Probability Plots **20**

B R code **23**

Bibliography **24**

List of Figures

2.1 Boxplots of $y_{i,j} - y_{i,\text{prev}}$ by state for the NBA. 9

2.2 Boxplots of $y_{i,j} - y_{i,\text{prev}}$ by state for the NHL. 10

3.1 The flow of 16 teams' playoff series from conference quarterfinals through NBA championship in 2011. 13

A.1 Probability plot of NBA model B. 20

A.2 Probability plot of NBA model C. 21

A.3 Probability plot of NHL model B. 22

List of Tables

2.1 The 16 possible states w_l in a best-of-seven playoff series with respect to the reference team. 8

2.2 The 24 parameters $d_{j,\text{prev}}$ in Model B and their simplified parameterization in terms of $d_1, \ldots, d_{15}$. 11

3.1 Parameter estimates, standard errors and diagnostics obtained from fitting Model A, Model B, Model C, Model D and Model E to the NBA data. An asterisk indicates significance at significance level 0.05. 14

3.2 Parameter estimates, standard errors and diagnostics obtained from fitting Model A, Model B, Model C, Model D and Model E to the NHL data. An asterisk indicates significance at significance levels 0.05. 16

Chapter 1

Introduction

1.1 What is Desperation in Sports?

The psychological effect of each player on a team, plays a big role during sporting events. Specially, this arises in time limited sports such as basketball, hockey, soccer, etc. Imagine that there are two teams playing in a series and the opponent is ahead by two or three games in a best of seven game series. Then the team of interest may feel a 'loss of hope' in the next game, or alternatively, the team may feel desperate to win the next game. For instance, Tim Thomas, the goalie of the Boston Bruins (who won the Conn Smythe Trophy as the most valuable player of the NHL playoffs, 2011) said after stopping 33 shots on May 23, 2011 'It was just reaction and, you know, desperation'.

1.2 Background of the Project

To investigate desperation, we consider playoff series in the National Basketball Association (NBA) and the National Hockey League (NHL) from 2003 through 2011. We analyse *spread* and *moneyline* data which were collected from the website www.covers.com. These *betting lines* (spreads and moneyline) are provided by sportsbooks. The reason for using betting lines instead of scoring data is that betting lines are less variable than scores from game to game or even series to series. Another advantage is that spreads provided by different sportsbooks for a particular NBA game, may vary at most by only two points.

In the literature, there were discussions on various psychological effects in the NBA and the NHL. For instance, Entine and Small (2008) suggest that lack of rest for the road team is an important contributor to the home court advantage in the NBA.

The assumption of independent and identically distributed (iid) trials has been taken into account in many analyses of sporting events in the literature. But consistent with the goal of our project, we do not follow this assumption. Supportively, Stern (1998a) criticizes the iid assumption using game results from the NBA and NHL playoffs.

1.3 Organization of the Project

In chapter 2, we provide a review of various elements of sports gambling that is relevant to our data. In particular, we convert sportsbook spreads and moneyline odds from the NBA and the NHL into winning probabilities. In other words, we convert betting lines into variables measuring team strength at a hypothetical neutral site. And then we introduce five regression models that are used to analyze the effect of the match situation in terms of team strength. After applying the methodology to NBA and NHL playoff data, we do our data analysis in chapter 3. We display our results and analysis for the NBA and NHL separately. In chapter 4, we summarize our results and discuss a possible direction for future work.

Chapter 2

Model Development

2.1 An Explanation of Sports Betting Lines

There are different types of bets in sports. For instance, spread bets, moneyline bets, over/under bets, proposition bets, etc. As mentioned in section 1.2, this project concerns spread bets (one team is typically favored over another by the pointspread) and moneyline bets (wagers on the winner are based on payoffs).

For illustration, consider the fourth game of the 2011 NBA finals between the Dallas Mavericks and the Miami Heat which took place in Dallas on June 7, 2011. The betting line corresponding to the pointspread was given by

$$\begin{array}{lll} \text{Miami Heat} & +3 & (-110) \\ \text{Dallas Mavericks} & -3 & (-110) \end{array} \tag{2.1}$$

In the betting line (2.1), -3 indicates that if a gambler wagers on Dallas, Dallas needs to win the game by more than three points in order for the gambler to be a winner. If Dallas wins exactly by three points then no one wins or losses. If Dallas wins by less than three points or loses the game then the gambler will be a loser. Further, +3 indicates that if a gambler wagers on Miami, the gambler can win, even if Miami loses, as long as they lose by less than three points.

In (2.1), the numbers in the brackets are the payout odds on the game. Lets consider both pointspreads and payout odds together. The betting line (2.1) shows how much a gambler has to wager in order to win \$100. Furthermore, if a gambler wagers \$110 on

Dallas to win by more than three points (negative pointspread) and Dallas achieves it, then the gambler receives \$210 (the \$110 that the gambler initially wagered plus the \$100 won). If Dallas wins by exactly three points, this is referred to as a *push* and the gambler receives only his wager of \$110. If Dallas wins by one or two points, or loses the game, then the gambler loses the \$110 wager. On the other hand if a gambler wagers \$110 on Miami and Miami wins or if Dallas wins by one or two points, then the gambler receives \$210. If a push occurs, the gambler receives \$110 and if Dallas wins by more than three points the gambler loses the \$110 wager.

Note that there are variations of the situation described above. A gambler does not need to bet \$110, but may bet any amount not exceeding the limit imposed by the sportsbook, and the amount won/lost is then proportional to the amount wagered. Also, the pointspread does not need to be in an integer. Sometimes it can be a number such as 2.5 in which case there is no possibility of a push.

For the NHL playoff data, we prefer to work with *moneyline* data rather pointspread data as in NBA. Moneyline odds are also known as American odds.

For illustration, consider the first game of the 2011 NHL finals between the Vancouver Canucks and the Boston Bruins which took place in Vancouver on June 01, 2011. The moneyline was given by

$$\begin{array}{ll} \text{Boston Bruins} & +173 \\ \text{Vancouver Canucks} & -188 \end{array} \tag{2.2}$$

The positive odds on Boston indicate how much money is won on a wager of \$100 and the negative odds indicate how much is needed to wager to win \$100. Referring to (2.2), suppose that a gambler wagers \$188 on Vancouver. Then \$188 is returned along with a profit of \$100 if Vancouver wins, and the gambler loses the \$188 wager if Vancouver loses. On the other hand, consider a gambler who wagers \$100 on Boston, and note that the American odds of +173 has a positive sign. In this case, the gambler receives \$100 back along with a profit of \$173 if the Bruins win, and the gambler loses the \$100 wager if the Bruins lose. In (2.2), Vancouver is the favorite and the American odds of +173 is provided to lure wagers towards Boston.

It is also possible to have moneylines where both odds are negative. This happens when teams are more evenly matched. Consider the fourth game of the 2011 NHL conference finals between Vancouver Canucks and the San Jose Sharks which took place in San Jose on May 22, 2011. The moneyline was given by

$$\begin{array}{ll} \text{Vancouver Canucks} & -102 \\ \text{San Jose Sharks} & -106 \end{array} \tag{2.3}$$

Here, it would require a wager of \$106 to win \$100 on San Jose, and a \$102 wager to win \$100 on Vancouver. In (2.3), San Jose is the favored team and Vancouver is the underdog team.

2.2 Converting Betting Lines to Probabilities

For the NBA, we wish to convert betting lines involving pointspreads to variables denoting team strength on a hypothetical neutral court. Let μ denote the pointspread for a team of interest where negative values indicate that the team is favored to win. Stern and Mock (1998) suggest that the point differential by which the team of interest defeats its opponent in US college basketball is well approximated by the normal(-μ, σ^2) distribution. There are various suggestions regarding the value of σ in literature. For instance, Entine and Small(2008) reported $\sigma = 11.1$, Gibbs(2007) suggested $\sigma = 11.4$ by using NBA data over the period 1993-2007 and Larsen, Price and Wolfers(2008) fit a normal distribution where $\sigma = 11.6$ is obtained (personal communication). We used $\sigma = 11.1$ for our analysis. The original pointspreads from the website take into account the home court advantage. Hence we need to eliminate it from our original data. Gandar, Zuber and Lamb (2001) suggest, for an NBA regular season game, the home court advantage $h \approx 4.0$ points. However we choose $h = 3.4$ for our analysis. Our rationale is based on the lack of balance in the regular season in the sense that visiting teams more often play back-to-back games. In the playoffs, both teams are equally rested. Additionally $h = 3.4$ is consistent with Entine and Small (2008) who analyzed the 2004-2005 and 2005-2006 NBA seasons, and gave a confidence interval (2.46,3.40) for h corresponding to the home court advantage for equally rested teams. However, our results in section 3.1 are not sensitive to small changes in the home court advantage parameter h and the normal standard deviation parameter σ.

We define the measure of strength y, for an NBA team with pointspread μ as

$$y = \ln(p/(1-p)) \tag{2.4}$$

where

$$p = \begin{cases} \Phi((-\mu - h)/\sigma) & \text{home game for the team of interest} \\ \Phi((-\mu + h)/\sigma) & \text{away game for the team of interest} \end{cases} \tag{2.5}$$

is the probability of victory for the team of interest on a hypothetical neutral court and Φ is the distribution function of the standard normal distribution. Note that the variable y in (2.4) has been transformed to the real line. Positive (negative) values of y indicate that the team of interest is stronger (weaker) than its opponent.

For the NHL, lets return to the Vancouver/Boston match and suppose that we wish to convert moneyline data to win probabilities on hypothetical neutral ice. Ignoring the sign in (2.2), the difference between +173 and -188 represents the vigorish. If there were no vigorish, then one might see a moneyline such as Vancouver -180.5 and Boston +180.5, in which case there would be no systematic advantage for the sportsbook and 180.5 would represent the "true" moneyline. To convert the moneyline data to win probabilities, we eliminate the vigorish by taking the midpoint between the odds. For example, in (2.2), we let p be the probability that Vancouver wins, and we set the expected profit from a \$180.5 wager on Vancouver equal to zero;

$$0 = 100p - 180.5(1-p)$$

which gives $p = 0.643$. Note that we obtain the same result by considering a wager on Boston and setting the expected profit equal to zero.

Now we need to eliminate home ice advantage from the winning probabilities. For that we use home ice winning percentage which is calculated by using NHL regular season data from 2006 through 2011. In this data, the home team has won 3353 games out of 6150 games. As a percentage, it is 54.5%. The home ice winning percentage 54.5% is a much smaller effect, compared to the home team wins (60.5%) in the NBA (Stefani 2008). Then to eliminate home ice advantage, we substract 0.045 from the calculated winning probability p if the team of interest is playing at home, and add 0.045 to the calculated winning probability p if the team of interest is playing on the road. For the above example, Vancouver is the home team and therefore its winning percentage on hypothetical neutral ice is $0.643 - 0.045 = 0.598$. And using (2.4), our measure of strength for Vancouver on hypothetical neutral ice is $y = \ln(0.598/0.402) = 0.397$.

2.3 Statistical Models

For each NBA and NHL team, we define the measure of strength y, as in equation (2.4). The variable y is our response variable. Our goal is to investigate how the team's strength varies during the series. For instance, can we see any difference in a team's strength, when it plays game1, game2, game3, etc in a series? Does the desperation take place in there? However, in our project, we believe that the strength of the team is a combination of its 'form' and its 'desperation'. 'Form' is the natural effect (inherent ability) which takes into account players' injuries, players' leave due to personal reasons, etc. Note that form can change from game to game. 'Desperation' is the psychological effect. Now we have an idea of covariates for the response variable.

Secondly, we need to examine our data format. Both NBA and NHL teams have the same home/away pattern, $HHAAHAH$ in a series (except NBA championship final series). H denotes the team of reference plays at home and A denotes the team of interest plays away (on the road). Also both NBA and NHL game series are best-of-seven series. However if a team wins four games consecutively, then the series will end with only four games. We do not include the NBA championship final series in our models since it has a different home/away pattern. Also we define the 'team of reference' as the team whose first game is at home. We use the following notation for the game 'state' of the series. Let w_l be the state of the series where w denotes the number of games won by reference team and l denotes the number of games lost by reference team. Therefore we have 16 possible states as given in table 2.1.

Note that the states for a series can be different in different series. We define the response variable is $y_{i,j} \in (-\infty, \infty)$ as the reference team's strength on neutral court in state j of series i.

State	w - l	State	w - l	State	w - l	State	w - l
0	0 - 0	4	0 - 1	8	0 - 2	12	0 - 3
1	1 - 0	5	1 - 1	9	1 - 2	13	1 - 3
2	2 - 0	6	2 - 1	10	2 - 2	14	2 - 3
3	3 - 0	7	3 - 1	11	3 - 2	15	3 - 3

Table 2.1: The 16 possible states w - l in a best-of-seven playoff series with respect to the reference team.

Now we can ask the same question we mentioned at the beginning of section 2.3, but in a different way. Can we see any difference in team strength, when a team plays in states 3 (3 - 0), 15 (3 - 3) and 12 (0 - 3)? Does desperation exist in some states?

We define our initial model,

$$y_{i,j} = f_{i,j} + s_j + \epsilon_{i,j} \tag{2.6}$$

where $f_{i,j}$ is the form of the team of reference in state j of series i, and s_j is the desperation in state j. The $\epsilon_{i,j}$ are error terms.

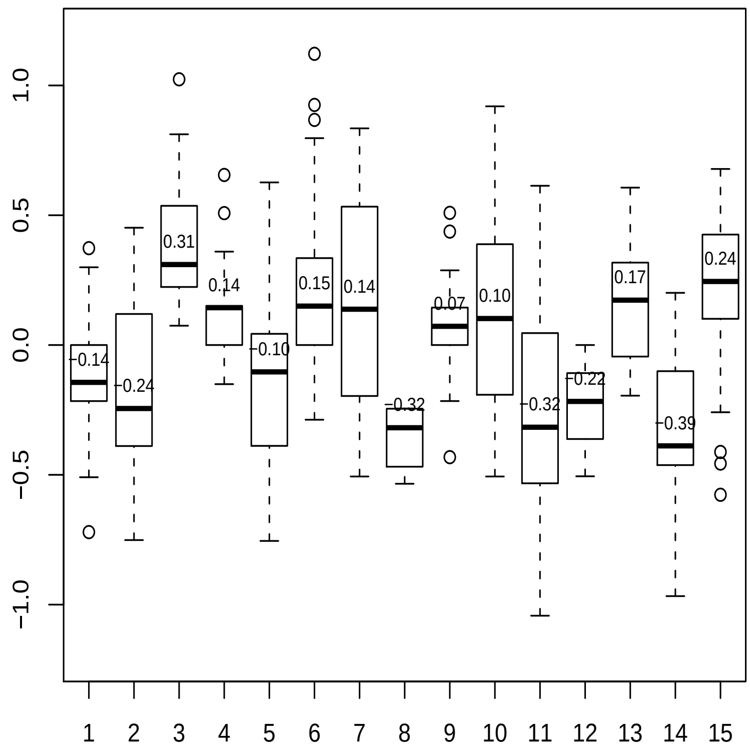
Our project goal is to explore the desperation variable s_j. However it is complicated to separate the form variable and the desperation variable in the model. Therefore to eliminate the form variable of our initial model, we make the following assumption. We assume that the form in the first game of a particular series (i.e. $f_{i,0}$) does not change throughout the series. Then using the assumption and the equation (2.6), we obtain

$$\begin{aligned} y_{i,j} - y_{i,0} &= (f_{i,0} + s_j + \epsilon_{i,j}) - (f_{i,0} + s_0 + \epsilon_{i,0}) \\ &= (s_j - s_0) + (\epsilon_{i,j} - \epsilon_{i,0}). \end{aligned} \tag{2.7}$$

For a second model, we assume that the form $f_{i,j}$ in a game j is equal to the form $f_{i,prev}$ in the previous game. We then obtain

$$\begin{aligned} y_{i,j} - y_{i,\text{prev}} &= (f_{i,\text{prev}} + s_j + \epsilon_{i,j}) - (f_{i,\text{prev}} + s_{\text{prev}} + \epsilon_{i,\text{prev}}) \\ &= (s_j - s_{\text{prev}}) + (\epsilon_{i,j} - \epsilon_{i,\text{prev}}). \end{aligned} \tag{2.8}$$

Figure 2.1: Boxplots of $y_{i,j} - y_{i,\text{prev}}$ by state for the NBA.



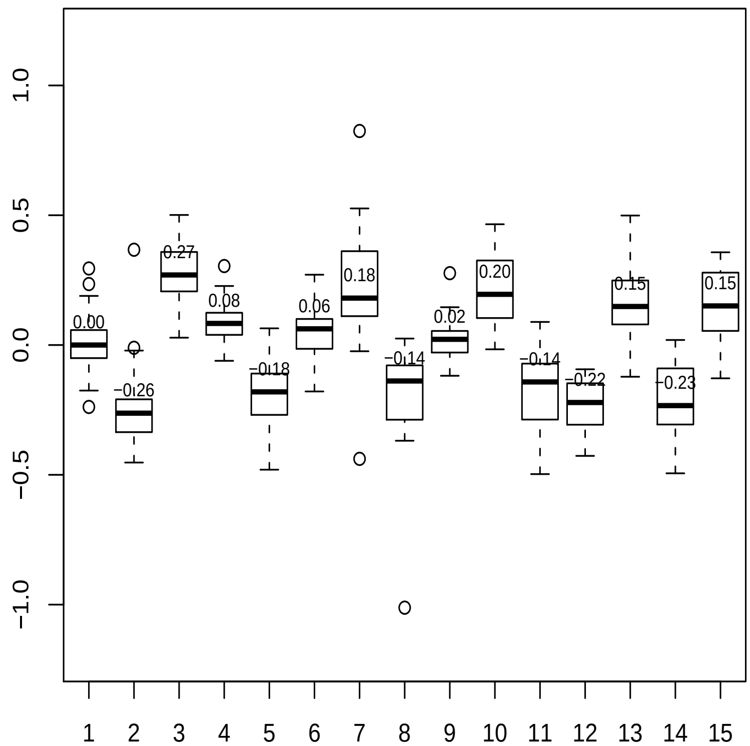
For both the NBA and NHL, boxplots of $y_{i,j} - y_{i,\text{prev}}$ with respect to the state are given in figure 2.1 and figure 2.2. The median values for each state are also display in the figures. The boxplots suggest that the states of a playoff series do have an effect.

Referring back to equation (2.7), we define model A,

$$y_{i,j} - y_{i,0} = d_j + \epsilon^*_{i,j} \tag{2.9}$$

where $d_j = s_j - s_0$ represents the change in desperation from the beginning of the series to state j, and the $\epsilon^*_{i,j} = (\epsilon_{i,j} + \epsilon_{i,0})$ are error terms.

Figure 2.2: Boxplots of $y_{i,j} - y_{i,\text{prev}}$ by state for the NHL.



From equation (2.8) we define model B,

$$y_{i,j} - y_{i,\text{prev}} = d_{j,\text{prev}} + \epsilon^*_{i,j} \tag{2.10}$$

where $d_{j,\text{prev}} = s_j - s_{\text{prev}}$ represents the change in desperation from the previous state to state j, and the $\epsilon^*_{i,j} = (\epsilon_{i,j} + \epsilon_{i,prev})$ are error terms.

The 24 combinations of $d_{j,\text{prev}}$ are shown in table 2.2. This allows us to parameterize model B in terms of the parameters in model A.

Further we were concerned whether the reference team's previous game outcome may

$d_{j,\text{prev}}$	Coding	$d_{j,\text{prev}}$	Coding	$d_{j,\text{prev}}$	Coding
$d_{1,0}$	d_1	$d_{7,3}$	$d_7 - d_3$	$d_{11,10}$	$d_{11} - d_{10}$
$d_{2,1}$	$d_2 - d_1$	$d_{7,6}$	$d_7 - d_6$	$d_{12,8}$	$d_{12} - d_8$
$d_{3,2}$	$d_3 - d_2$	$d_{8,4}$	$d_8 - d_4$	$d_{13,9}$	$d_{13} - d_9$
$d_{4,0}$	d_4	$d_{9,5}$	$d_9 - d_5$	$d_{13,12}$	$d_{13} - d_{12}$
$d_{5,1}$	$d_5 - d_1$	$d_{9,8}$	$d_9 - d_8$	$d_{14,10}$	$d_{14} - d_{10}$
$d_{5,4}$	$d_5 - d_4$	$d_{10,6}$	$d_{10} - d_6$	$d_{14,13}$	$d_{14} - d_{13}$
$d_{6,2}$	$d_6 - d_2$	$d_{10,9}$	$d_{10} - d_9$	$d_{15,11}$	$d_{15} - d_{11}$
$d_{6,5}$	$d_6 - d_5$	$d_{11,7}$	$d_{11} - d_7$	$d_{15,14}$	$d_{15} - d_{14}$

Table 2.2: The 24 parameters $d_{j,\text{prev}}$ in Model B and their simplified parameterization in terms of $d_1, \ldots, d_{15}$.

affect its current game. For that we introduced another variable I_w such that,

$$I_w = \begin{cases} 1 & \text{if reference team won previous game} \\ 0 & \text{if reference team lost previous game} \end{cases}$$

Hence, by adding I_w to equation (2.8) we define model C,

$$y_{i,j} - y_{i,\text{prev}} = d_{j,\text{prev}} + \delta I_w + \epsilon^*_{i,j} \tag{2.11}$$

where $d_{j,\text{prev}} = s_j - s_{\text{prev}}$ represents the change in the desperation from the previous state to state j, and the $\epsilon^*_{i,j}$ are error terms.

Also we were interested in modifications of model C. For example, it is possible that the win/loss of the previous game has a greater/lesser effect early in the series. We therefore defined model D. In model D, we gave more weights on I_w in last games in a series.

$$y_{i,j} - y_{i,\text{prev}} = d_{j,\text{prev}} + \delta I_w \times [K(n) - 1] + \epsilon^*_{i,j} \tag{2.12}$$

where $K(n)$ represents the series game number; $K(n) = 2, ..., 7$.

And also by investigating the boxplots closely, we noticed that game four in a series is a landmark game. Therefore, we defined model E,

$$y_{i,j} - y_{i,\text{prev}} = d_{j,\text{prev}} + \delta I_w \times \begin{cases} 1 + |K(n) - 2.9| & if\ K(n) \leq 4 \\ 1 + |K(n) - 5.9| & if\ K(n) > 4 \end{cases} + \epsilon^*_{i,j} \tag{2.13}$$

Chapter 3

Data Analysis

3.1 NBA Data

From the regular season NBA matches, the best 16 teams are eligible for the playoffs in each year. These 16 teams are the teams who play for the NBA championship. Figure 3.1 shows the flow of teams' playoffs from conference quarterfinals through NBA championship for the year 2011.

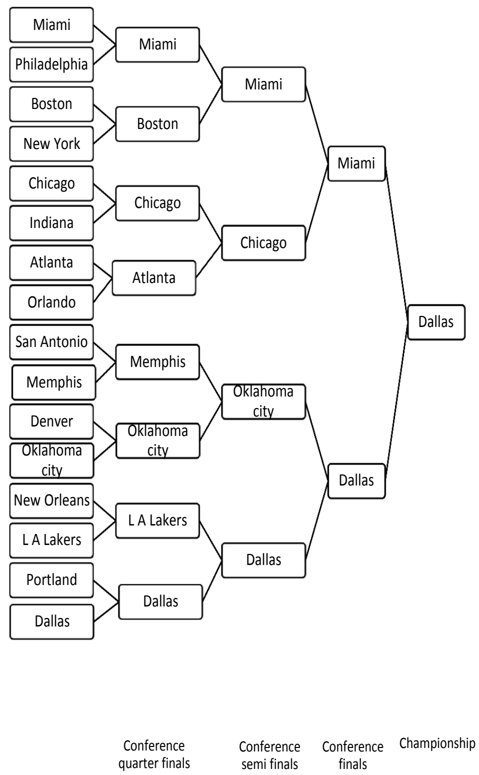
As we mentioned in section 2.3, we do not include the final championship series in each year since it has a different home/away pattern. Therefore our data consists 14 series for each year, and 126 series from 2003 to 2011. In our data, the majority of matches belong to state 1 (1 - 0), state 2 (2 - 0) and state 6 (2 - 1). And state 12 (0 - 3) has the lowest number of matches.

The summary of the five fitted models which were introduced in section 2.3 is given in table 3.1. In table 3.1, the R^2 diagnostic tells us that model A does not fit as well as the other four models. This illustrates that the models which use the assumption that the form in a game is equal to the form in the previous game is preferable. The other four models are all comparable in terms of fit. In view of model simplicity and the number of significant parameters contains in a model, we prefer model C. Normal probability plots of models B and C are given in appendix A, figure A.1 and figure A.2.

Comparing all 15 parameter estimates in model C, d_3, d_6, d_7 and d_{15} have large positive estimates and d_8, d_{12} and d_{14} have large negative estimates.

The parameter d_3 corresponds to state 3 (3 - 0), where the reference team is ahead of the opponent by 3 games to zero. In this case, the reference team needs one more game to win

Figure 3.1: The flow of 16 teams' playoff series from conference quarterfinals through NBA championship in 2011.



the series. Here, δ gives a negative effect but is small related to the estimate of d_3. However, the reference team's feeling of 'one game' overcomes in this situation. Consequently for the opponent, game 4 is a fateful event. In NBA history we can see, if a team reaches the 3 – 0 state then the opponent has never won the series. Also, a similar effect happens in state 7 (3 – 1). Even if the opponent has won one game, the reference team is confident of winning the series. According to the 2003 through 2011 NBA playoffs, if a reference team reaches 3 – 1 state then it will win the series 100% of the time. From the opponents point of view, they give up on the series after the reference team reaches the states 3 or 7. In state 6 (2 – 1) the reference team tries to reach 3 – 1 which is very stable position as described above.

Parameter	Model A	Model B	Model C	Model D	Model E
d_1	-0.10* (0.03)	-0.10* (0.03)	0.01 (0.05)	-0.08* (0.03)	0.01 (0.05)
d_2	-0.26* (0.04)	-0.15* (0.03)	-0.05 (0.05)	-0.10* (0.04)	-0.09* (0.04)
d_3	0.21* (0.06)	0.39* (0.06)	0.50* (0.07)	0.47* (0.06)	0.52* (0.07)
d_4	0.12* (0.05)	0.12* (0.05)	0.12* (0.05)	0.12* (0.05)	0.12* (0.05)
d_5	-0.12* (0.04)	-0.16* (0.04)	-0.10* (0.05)	-0.13* (0.04)	-0.12* (0.04)
d_6	-0.04 (0.04)	0.19* (0.03)	0.23* (0.04)	0.22* (0.04)	0.24* (0.04)
d_7	0.14* (0.05)	0.16* (0.05)	0.25* (0.06)	0.24* (0.06)	0.25* (0.06)
d_8	-0.16 (0.12)	-0.35* (0.12)	-0.36* (0.12)	-0.36* (0.12)	-0.35* (0.12)
d_9	-0.07 (0.06)	0.07 (0.06)	0.08 (0.06)	0.08 (0.06)	0.08 (0.06)
d_{10}	0.06 (0.04)	0.10* (0.04)	0.12* (0.04)	0.12* (0.04)	0.12* (0.04)
d_{11}	-0.15* (0.04)	-0.24* (0.04)	-0.14* (0.05)	-0.13* (0.06)	-0.18* (0.05)
d_{12}	-0.32* (0.15)	-0.23 (0.14)	-0.23 (0.14)	-0.23 (0.14)	-0.23 (0.14)
d_{13}	-0.00 (0.07)	0.16* (0.07)	0.17* (0.07)	0.17* (0.07)	0.17* (0.07)
d_{14}	-0.31* (0.06)	-0.32* (0.06)	-0.25* (0.06)	-0.24* (0.06)	-0.28* (0.06)
d_{15}	-0.01 (0.06)	0.19* (0.06)	0.22* (0.06)	0.24* (0.06)	0.23* (0.06)
δ			-0.11* (0.04)	-0.02* (0.01)	-0.06* (0.02)
R^2	0.20	0.30	0.31	0.31	0.31

Table 3.1: Parameter estimates, standard errors and diagnostics obtained from fitting Model A, Model B, Model C, Model D and Model E to the NBA data. An asterisk indicates significance at significance level 0.05.

In state 15 (3 _ 3), both teams have won same number of games and in addition, both teams have one more game to win the series. Recall that the home/away pattern in NBA playoffs ($HHAAHAH$). Therefore the home court advantage takes place to increase the strength of the reference team.

Estimates for d_8 state 8 (0 _ 2) and d_{12} state 12 (0 _ 3) have large negative values, which implies the reference team gives up hope to win game 3 and game 4 on the road. In state 14 (2 _ 3), the reference team is despairing since the opponent has won 3 out of 5 games and the opponent has to win one more to win the series. Not only that but game 6 takes place on the road.

Finally the negative δ term implies that if the reference team won the previous game then they tend to enjoy the next game without having any stress. In model C, the estimate of δ is -0.11 and in models D and E, δ shows a minor effect. On the other hand, it is possible that the I_w parameter is a psychological effect.

Further, if we do consider the I_w effect as a psychological effect, we can stick to model B.

For the illustration of model B, we concern only parameter(s) which satisfy $|estimate(s)| > 0.2$. The rationale behind this is $|d_j| > 0.2$ corresponds to a 0.05 increase in probability from probability p. Note that $d_j = (s_j - s_0)$ and hence, it equals to $\ln(p_j/1-p_j) - \ln(p_0/1-p_0)$. In model B, only state 3 (3 _ 0) satisfies $d_j > 0.2$.

Estimates for d_8, d_{11}, d_{12} and d_{14} have large negative values. In state 8 (0 _ 2) and d_{12} state 12 (0 _ 3), the reference team has lost their first two home games. The next two games will take place on the road and there is little hope for the reference team. In state 11 (3 _ 2), the reference team has won 3 out of 5 games and game 6 is on the road. The home court advantage takes place in this situation. The reference team is convinced they will win game 7 which is at home and chill out in game 6. Consequently, the opponent tries its best to win game 6. In state 14 (2 _ 3), the reference team has won only two games even though they played 3 games at home. And the next game is game 6 which is on the road. The reference team feels less confident in game 6 and loses hope.

3.2 NHL Data

From the regular season NHL matches, the best 16 teams are eligible for the playoffs in each year. These 16 teams are the teams who play for the NHL Stanley Cup championship.

The NHL data consists of 15 series for each year (the final championship series also takes place since it has the same home/away pattern as the other series). There are 90 series from 2006 to 2011. In our data, the majority of matches belong to state 1 (1 _ 0), state 6 (2 _ 1) state 2 (2 _ 0) and state 5 (1 _ 1). And, as in the NBA, state 12 (0 _ 3) has the lowest number of matches. Our NHL data consists of 514 total number of matches. The summary of the five fitted models which was introduced in section 2.3 is given in table 3.2.

In table 3.2, the R^2 fit diagnostic indicates that the model A does not fit as well as the other four models. This illustrates that the models which use the assumption that the form in a game is equal to the form in the previous game, fit well. These four models B, C, D and E have the same $R^2 = 0.62$ and a minor effect corresponding to δ (although it is possible that the I_w parameter is a psychological effect). Therefore by considering the simplicity of the models, we prefer model B. Normal probability plot of model B is given in appendix A, figure A.3 and the Rcode of the data analysis is given in appendix B.

Parameter	Model A	Model B	Model C	Model D	Model E
d_1	0.10 (0.02)	0.01 (0.02)	-0.01 (0.02)	-0.01 (0.02)	-0.01 (0.03)
d_2	-0.22* (0.02)	-0.25* (0.02)	-0.27* (0.03)	-0.27* (0.02)	-0.26* (0.02)
d_3	0.12* (0.03)	0.27* (0.03)	0.25* (0.04)	0.24* (0.03)	0.24* (0.04)
d_4	0.09* (0.03)	0.09* (0.02)	0.09* (0.02)	0.09* (0.02)	0.09* (0.02)
d_5	-0.16* (0.02)	-0.19* (0.02)	-0.20* (0.02)	-0.20* (0.02)	-0.19* (0.02)
d_6	-0.16* (0.02)	0.04* (0.02)	0.03 (0.02)	0.03 (0.02)	0.03 (0.02)
d_7	0.14* (0.03)	0.22* (0.02)	0.21* (0.03)	0.19* (0.03)	0.20* (0.03)
d_8	-0.14* (0.04)	-0.23* (0.04)	-0.23* (0.04)	-0.23* (0.04)	-0.23* (0.04)
d_9	-0.13* (0.03)	0.02 (0.03)	0.02 (0.03)	0.02 (0.03)	0.02 (0.03)
d_{10}	0.03 (0.03)	0.21* (0.02)	0.20* (0.02)	0.19* (0.02)	0.20* (0.02)
d_{11}	-0.10* (0.02)	-0.18* (0.02)	-0.19* (0.02)	-0.21* (0.03)	-0.19* (0.02)
d_{12}	-0.43* (0.05)	-0.23* (0.05)	-0.23* (0.05)	-0.23* (0.05)	-0.23* (0.05)
d_{13}	-0.04 (0.04)	0.17* (0.03)	0.16* (0.03)	0.16* (0.03)	0.16* (0.03)
d_{14}	-0.23* (0.04)	-0.22* (0.03)	-0.23* (0.03)	-0.24* (0.03)	-0.23* (0.03)
d_{15}	0.01 (0.03)	0.15* (0.03)	0.15* (0.03)	0.14* (0.03)	0.14* (0.03)
δ			0.02 (0.02)	0.01* (0.00)	0.01 (0.01)
R^2	0.48	0.62	0.62	0.62	0.62

Table 3.2: Parameter estimates, standard errors and diagnostics obtained from fitting Model A, Model B, Model C, Model D and Model E to the NHL data. An asterisk indicates significance at significance levels 0.05.

As in the NBA data analysis, first we look at larger positive and negative estimates in the model. In other words, we consider the states which have larger effects. Parameters d_3, d_7 and d_{10} have large positive estimates while d_2, d_8, d_{12} and d_{14} have large negative values in model B. The large positive estimate for state 3 (3 _ 0) illustrates that the reference team's psychological advantage is very high in game 4 after they have won the first three games. In other words the reference team does not show any desperation in game 4. The positive estimate of d_7 state 7 (3 _ 1) illustrates, after the reference team won three games including a win of on the road game, they are more confident on game 5 which takes place at home. It is clear that the home ice advantage effects to the strength of a team. We can see this in d_{10} state 10 (2 _ 2), where both teams have won an equal number of games. For the reference team, the next game (game five) is at home and they do have a hope to win the game.

In state 2 (2 _ 0), the reference team has won the first two games, and they may relax in the third game since it is on the road. On the other hand, home ice advantage arises in this situation. Hence the opponent is more powerful than the reference team. However the opposite of this happens in d_8 state 8 (0 – 2). The reference team has lost their first two home games and has little hope in game 3 since it takes place on the road. A similar thing happens in state 12 (0 – 3). The reference team has lost the first three games and despairs to win game 4 on the road. For state 14 (2 – 3), the reference team has won only two games even though they played 3 games at home. And game six is on the road. Therefore the reference team feels less confident in game six. In other words, the reference team is desperate to win game six.

In table 3.2, we can see that the estimates and the standard deviations of d_4, d_8, d_9 and d_{12} are the same in all four models B, C, D and E. Further, the estimate of d_4 is same in all five models.

Chapter 4

Conclusions

4.1 Discussion

In each state, the variation of strength from the previous state to that state (i.e. $y_{i,j} - y_{i,\text{prev}}$) is larger in the NBA than the NHL. And also for all states, the strength from the previous state to the current state in the NHL, is more concentrated around zero than in the NBA (figure 2.1 and figure 2.2). As a result we obtained estimates with smaller standard errors in the NHL than in the NBA and also we see that the NHL team's strength is less affected by the state. We defined our initial model by considering 'form' and 'desperation'. However, the models which use the assumption that the form in a game is equal to the form in the previous game, give better results. Further, the additional parameter (I_w) which quantifies the effect of the previous game win/loss, was considered as a psychological effect in both the NBA and the NHL. Hence for both NBA and NHL, the most suitable model is model B which combines the change in strength and change in desperation from the previous state to a state.

Our goal is to investigate how a team's strength varies from game to game. In other words, we explore the desperation variable. We use $|estimate(s)| > 0.2$ as a cut off mark to filter out small estimates. We obtained estimates of d_3, d_7, d_8, d_{10}, d_{11}, d_{12} and d_{14} as large estimates in the NBA and the NHL. However, the largest effects can be found when one team has won zero games (states 0 _ 2, 2 _ 0, 0 _ 3, 3 _ 0), except game 1. Not only that, it is clear in the NBA, the largest negative effects can be found when a team is close to elimination (state 2 _ 3). Further, for both NBA and NHL, we can see the reference team's desperate situations in state 8, state 12 and state 14. However, as a result of home court advantage, the NBA reference team is desperate to win game 6 in states 11 or 14. Similarly, as a result of home ice advantage, the NHL reference team hopes to win game 5 in state 7 or state 10 and is desperate to win game 6 in state 14. The home team advantage effects desperation in both NBA and NHL sports.

The effect of the previous games win/loss affects negatively in the NBA such as if they win the previous game then they may relax in the subsequent game. Consequently, the effect of the previous game win/loss affects positively in the NHL. However the affect is very small and ignorable.

4.2 Future Work

It sometimes happens that some of the observations used in a regression analysis are less reliable than others. Hence, the variances of the observations may not all equal or on the other hand the observations may be correlated. In our data, the variables $y_{i,j} - y_{i,\text{prev}}$ may be correlated. If this is the case, it is worthy to find weighted least squares (or generalized least squares) estimates instead of simple regression model estimates. For instance, we can obtain weighted least squares estimator $\hat{\beta}$, by using the following equation

$$\hat{\beta} = (X^{'}V^{-1}X)^{-1}X^{'}V^{-1}Y.$$

Here, $\hat{\beta}$ are the estimates of the d_js, X is the design matrix for d_j, Y is the response variable $(y_{i,j} - y_{i,\text{prev}})$ vector and V is a known matrix (can be found using data).

Appendix A

Probability Plots

Figure A.1: Probability plot of NBA model B.

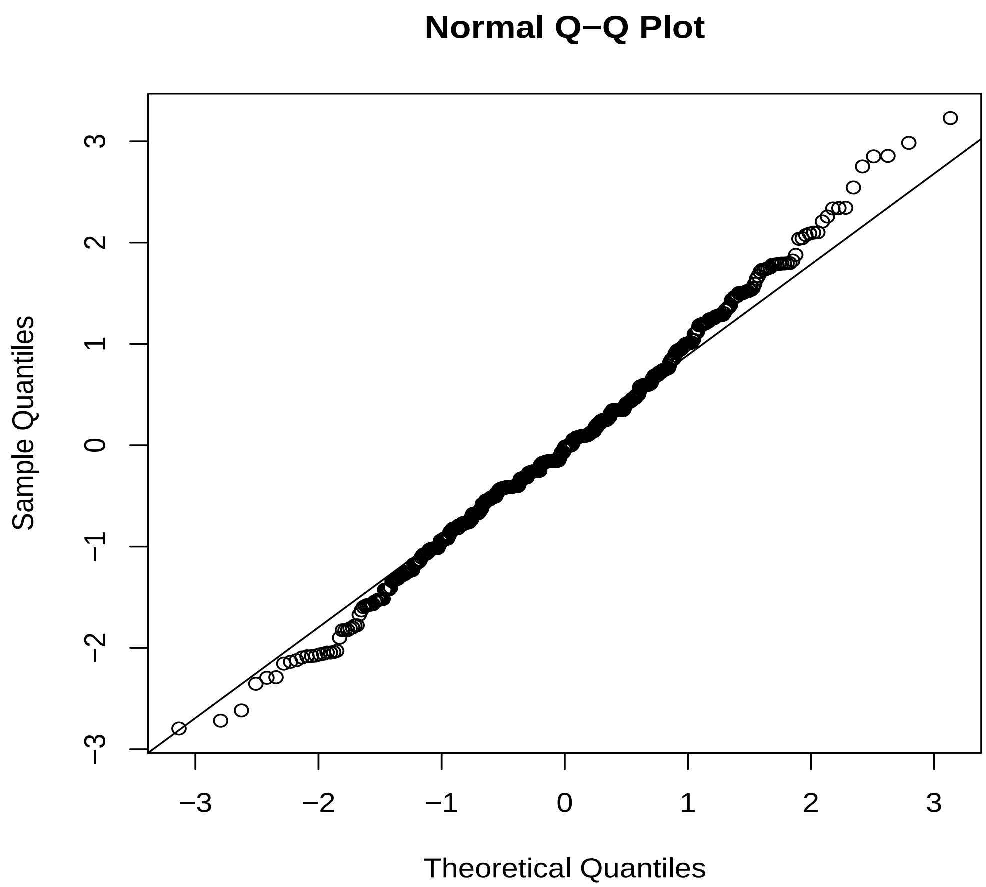


Figure A.2: Probability plot of NBA model C.

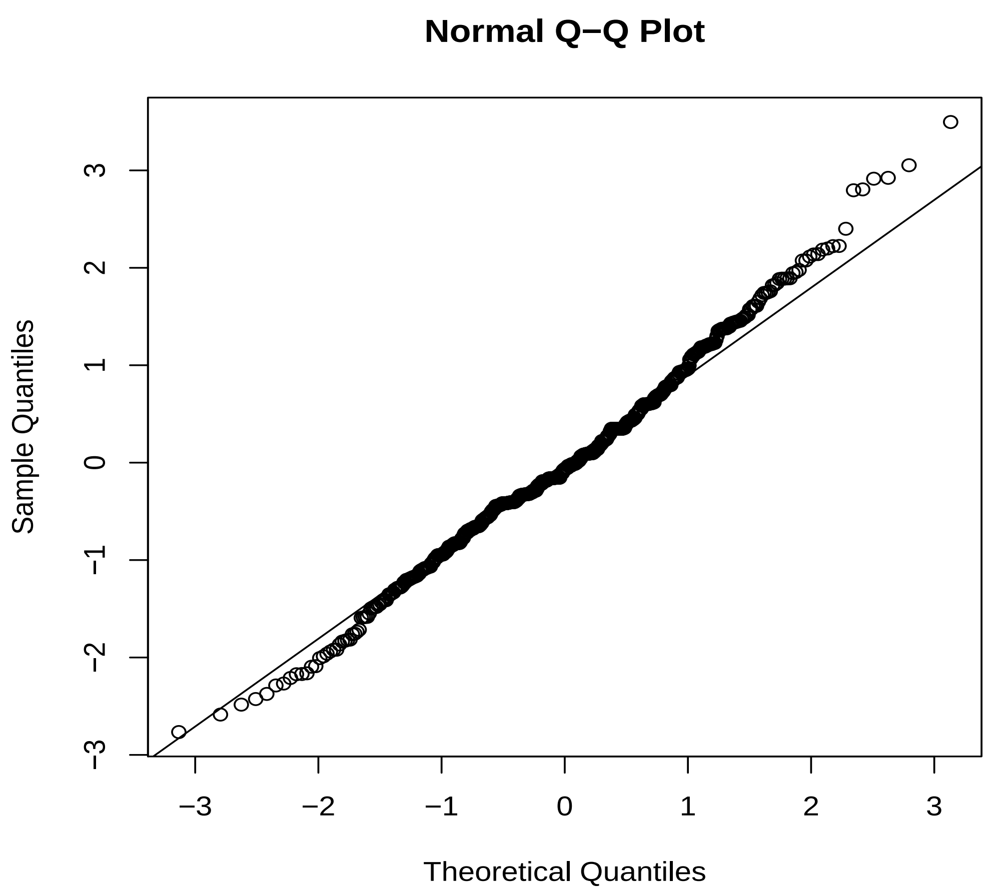
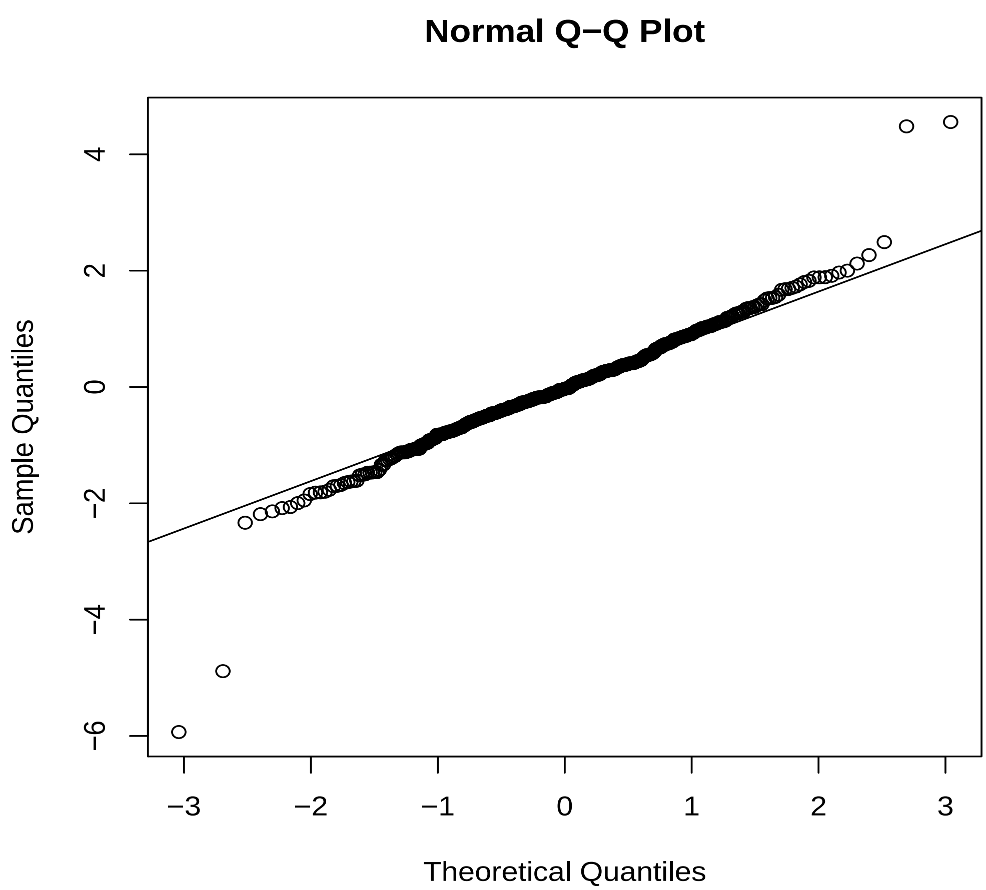


Figure A.3: Probability plot of NHL model B.



Appendix B

R code

```
# Yj.Y0     = the change in strength from the begining of the series to state j.

# Yj.Yprev  = the change in strength from the previous state to state j.

# Situation = the state.

# Iwin      = 1 if reference team won previous game, otherwise 0.

# IwinKi    = Iwin * |K(n)-1| ; K(n) represents the series game number.

# IwinKiE   = Iwin * |K(n)-2.9| if K(n) less than or equal to 4,
              otherwise Iwin * |K(n)-5.9|.


### Boxplots (yj-y0) and (yj-ypre) vs Situation

boxplot(Yj.Y0 ~ Situation)
medians <- by(Yj.Y0v, Situation, median)
text(1:15 , medians,
     labels = formatC(medians, format = "f",
                      digits = 2),
     pos = 3, cex = 0.7, col = "black")

boxplot(Yj.Yprev ~ Situation, ylim = c(-1.2,1.2), xlim = c(1,15), cex.axis=0.9)
medians <- by(Yj.Yprev, Situation, median)
text(1:15 , medians,
     labels = formatC(medians, format = "f",
                      digits = 2),
     pos = 3, cex = 0.7, col = "black")
```

```
### Models A, B, C, D, E
lmA<-lm(Y.YOv ~ factor(Situation)-1)
summary(lmA)

### histogram for residuals
resnba<-resid(lmA)
hist(resnba)

### normal prob. plot for residuals
stdres<-rstandard(lmA)
qqnorm(stdres)
qqline(stdres)

lmB<-lm(Y.Yprev ~ factor(Situation)-1)

lmC<-lm(Y.Yprev ~ factor(Situation) + Iwin -1)

lmD<-lm(Y.Yprev ~ factor(Situation) + IwinKi -1)

lmE<-lm(Y.Yprev ~ factor(Situation) + IwinKiE -1)
```

Bibliography

[1] Entine, O.A. and Small, D.S. The role of rest in the NBA home-court advantage. *Journal of Quantitative Analysis in Sports*, 4(2): Article 6, 2008.

[2] Gandar, J.M., Zuber, R.A. and Lamb, R.P. The home field advantage revisted: a search for the bias in other sports betting markets. *Journal of Economics and Business*, 53(4): 439–453, 2001.

[3] Gibbs, J. An economic analysis of the National basketball Association's point spread betting market. *Undergraduate Honors Thesis, Department of Economics, Stanford University*, 2007.

[4] Stefani, R. Measurement and interpretation of home advantage. In Statistical Thinking in Sports (editors J. Albert and R.H. Koning). *Chapman & Hall/CRC*, pages 203–216, 2008.

[5] Stern, H.S. Best-of-seven playoff series. In the column, A Statistician Reads the Sports Pages. *Chance*, 11(2): 46–49, 1998.

[6] Stern, H.S. and Mock, B. College basketball upsets: will a 16-seed ever beat a 1-seed?. In the column, A Statistician Reads the Sports Pages. *Chance*, 11(1): 26–31, 1998.