Forecasting NBA Basketball Playoff Outcomes Using the Weighted Likelihood *

Feifang Hu[1] and James V Zidek[2]

[1] University of Virginia and [2]University of British Columbia

Abstract

Predicting the outcome of a future game between two sports teams poses a challenging problem of interest to statistical scientists as well as the general public. To be effective such prediction must exploit special contextual features of the game. In this paper, we confront such features and address the need to: (i) use all relevant sample information; (ii) reflect the home game advantage. To do so we use the weighted likelihood. Finally we demonstrate the value of the method by showing how it could have been used to predict the 96/97 NBA playoff results. Our weighted likelihood-based method proves to be quite accurate.

Key words: Likelihood; relevance weighted likelihood; predicting basketball scores; sports statistics; NBA; basketball.

*The work reported in this article was partially supported by a grant from the Natural Science and Engineering Research Council of Canada

1 Introduction

This paper demonstrates the use of the weighted likelihood (WL) to predict the winner of National Basketball Association (NBA) final games and applies the method in each of the games played in the 1996-97 finals between the two top ranked teams, the Chicago Bulls and the Utah Jazz. However, as we try to indicate, the WL has much wider applicability inside as well outside the domain of sports.

That domain has generated a lot of interest in statistics. For example, the March 13, 2003 edition of North Carolina's "The News & Observer" (page 6D) carries the column caption "Profs predict bubble teams". The article goes on to describe how two academics, Jay Coleman and Allen Lynch, in apparently unpublished work, produce a method using "SAS analytics" for predicting the 65 NCAA teams to appear in the 2003 NCAA men's basketball tournament. However,

> "Coleman said the formula isn't designed to make bolder predictions, such as forecasting the tournaments's winner,"

(according to the article), although the latter would have brought their work closer to the objectives of the present paper.

The abundance of data on the WWW (where we found the data for this paper) demonstrates that interest, as do the statistical methods found there. At http://www.rawbw.com/ deano/ we found articles on, and links to a variety of such methods. We learned that some of these methods are

> "similar to those found in the analysis of baseball, particularly those of Bill James, whose work popularized the study of statistics in sports because it was scientifically solid and entertaining to read."

A book has even been written about statistics in sports (Bennett 1998), and refereed papers published. Harville (1977) uses regression analysis to rate high school and college football teams (from first to last) based on observed score differences. The ratings are then based on the estimates of a random effects model parameter associated with each team. In a later paper (Harville 1980), he develops a method for forecasting the point spread of NFL games, again using regression methods, again based on historical point spreads, and linear random effects models. The topic of that paper comes closer to ours, since these forecasted point spreads can be used to predict NFL game winners.

In related papers, Schwertman et al (1996) and Carlin (1996) tackle NCAA basketball. Both papers (like this one) estimate the probability that team i beats j. They (unlike us) are based on pre-game information. The first uses a logistic regression analysis of win - loss records and various functions of seed numbers, as a way of incorporating prior knowledge and expert opinion. The second extends earlier unpublished work of Schwertman et al (1993) by using other external information such as "...the RPI index, Sagarin ratings, and so on..." in addition to seed numbers. Like Harville (1980) and Stern (1992), Carlin uses published point spreads to capture pregame information and does a linear regression analysis of observed point spreads on pregame information. Models derived from that analysis can be used to predict game winners.

Kuonen and Roehrl (2000) use a logistic model like that of Schwertman et al (1996) to explore the credibility of France's win in the 1998 World Cup of soccer tournament. The pre-game ratings had put Brazil well ahead of France casting doubt on the victory. They conclude France's victory should not have come as a surprise. (They also give a recent review of literature on the subject of predicting wins.)

Our approach unlike those described above, does not attempt to take pre-game information into consideration although it may be possible to do that through the weights in the WL. That issue remains to be explored. Instead, our goal is to introduce the WL method and show how it can be used. No doubt improvements that build on earlier work could enhance the method. However, we do assess our approach against a logistical method that embraces the celebrated method of Bradley and Terry (1952) that also underlies the work of Schwertman et al (1996).

The genesis of our work lies in two statistical problems encountered in sports: (i) the prediction of the outcome of a future game between two specified sports teams; (ii) the assessment of the accuracy of this prediction. Since typically these two teams will not have met more than just a few times in the given season, little direct information will be available to the forecaster. The consequent small sample size will make naive predictions inaccurate and the associated prediction intervals excessively large.

Turning to the NBA's 1996-97 finals, we note that the winner was the team that won the best of 7 games. To predict that outcome, one might sequentially determine the prediction probability of a Bulls win in each of a series of successive games. To find that probability, the 1996-97 season data would be used. However, the Bulls met the Jazz just twice, providing the only 'direct' information available, in the terminology of Hu and Zidek (1993) and Hu (1994). However, that small sample cannot generate accurate predictions.

To overcome this data deficiency, observe that the Bulls (like the Jazz) played 82 games in the season (2 with the Jazz and 80 with other teams). The 160 games these two played against other teams provide "relevant" information, in the Hu-Zidek terminology.

To use both the "direct" and "relevant" information in some simple yet flexible way, Hu (1994) proposes the "relevance weighted likelihood". Hu and Zidek (2002) extend that likelihood and Wang (2001) further extended it to get the "weighted likelihood (WL)", the terminology we use in this paper.

The method of weighted likelihood has been applied to a neurophysiology experiment (Hu and Rosenberger, 2000). In that paper, they find that both bias and mean square error are significantly reduced by using the weighted likelihood method. Hu and Zidek (2001) use the WL to predict the number of goals (with prediction intervals) for each of the Vancouver Canucks and Calgary Flames in their NHL games against each other during the 1996-97 season. They (Hu and Zidek 2002) show how the WL can be used to construct generalizations of the classical Shewhart control charts. Their generalization includes the moving average and exponentially moving average charts and allows for a variety of failure modes when processes go out of control. This application introduces the weighted likelihood ratio test. In that same paper, they show how the James Stein estimator, including generalizations, can be found with the WL.

A particularly important class of applications arise in estimating parameters that are interrelated, leading to natural relationships among the associated populations and inducing transfers of information from their associated samples. Van Eeden and Zidek (2002) show how such interrelations may be exploited through the WL when the means of two normal populations with known variances are ordered. The analogous problem when the mean difference is bounded is treated in Van Eeden and Zidek (2000). Finally, we would mention an application to disease mapping in Wang (2001).

In Section 2, we apply the WL in the NBA forecasting application above by taking advantage of special features of sports data. The maximum WL estimator (MWLE) is developed for predicting the result of a future game. The mean square error of this MWLE is given. Moreover, we construct approximate confidence intervals using the asymptotic theory for the MWLE given by Hu (1997).

In Section 3, we apply the method developed in Section 2 to predict the 1996-97 NBA playoff results, specifically for games involving the Chicago Bulls and the Utah Jazz. Our predictions agree quite well with the actual outcomes.

To validate that positive performance assessment, in Section 3 we consider the playoff games played by the Bulls against each of three other teams, the Miami Heat, the

Atlanta Hawks, and the New York Knicks. Similarly, playoff games between the Heat and Knicks are considered. These additional predictions are also in good agreement with the actual game outcomes.

Many other approaches can be taken in our application. In Section 4, our method is shown to compare favorably with a 'purpose built' competitor, an extension of the Bradley Terry model (Bradley and Terry 1952). Moreover, it proves to have all the flexibility and much of the simplicity of its classical predecessor proposed by Fisher. Thus, we are able to recommend it as a practical alternative to its competitors for the application considered.

2 Sports Data and the WL.

2.1 Contextual features.

Usually in sports, the outcome of any one game derives from the combined efforts of two teams that have seldom played each other before. Yet these games yield the only direct sample information available about the relative strength of these two teams. At the same time, each of these teams will have played many games against other teams thereby generating relevant (although not direct) sample information. The predictive probability of a win in the next game between these two teams, should combine both kinds of information.

In some sports, the home team has a great advantage (see Section 3) that must be accounted for when the data are analyzed (although in their application, Hu and Zidek (2002) ignored that advantage). Finally, the outcome of any one game will depend on both the offensive and defensive capabilities of the teams involved. Information about those capabilities is revealed through the scores of games involving the teams and ideally, would be used to help predict the outcome of future games. Hu and

Zidek (2001) use the weighted likelihood for that purpose in another sports application. However, in this paper we do not assume those scores to be available and more simply base our methodology on the binary outcomes of those games instead.

2.2 The Weighted Likelihood.

To develop a statistical model for the analysis of sports data, one should recognize the distinctive contextual features described in the last subsection. Let $Y_{AB}(h)$ be a Bernoulli random variable that is 1 or 0 according as the home team, A, wins against team B or not. Similarly, let $Y_{AB}(r)$ be a random variable that is 1 or 0 according as team A wins against team B when team B is at home. Note that $Y_{AB}(h) = 1 - Y_{BA}(r)$. As an approximation, assume the time series of Y's for different games and team pairs are independent in this paper. Clearly, a more sophisticated approach like that of Hu, Rosenberger, and Zidek (2000) would allow dependent game outcomes.

Suppose the $\{Y_{AB}(h)\}$ and $\{Y_{AB}(r)\}$ have probability density functions $f(y, p_{AB}(h))$ and $f(y, p_{AB}(r))$ respectively. To predict the game result, $(Y_{AB}(h), Y_{BA}(r))$ or $(Y_{AB}(r), Y_{BA}(h))$, we have to estimate the parameters $p_{AB}(h)$ and $p_{AB}(r)$.

To create the weights required in implementing the WL, we choose the same weight in the likelihood factor corresponding to each of the games A played against teams other than B, irrespective of the opponent. From Hu and Zidek (2002), we may use the weighted likelihood method to estimate the parameters $p_{AB}(h)$ and $p_{AB}(r)$. The log weighted likelihood of $p_{AB}(h)$ thus becomes

$$\begin{aligned}\sum_{i=1}^{k_{AB}} \log f(y_{AB}(h), p_{AB}(h)) + \alpha_{AB}(h) \sum_{A(B)} \log f(y_{A(B)}(h), p_{AB}(h)) \\ + \beta_{AB}(h) \sum_{(A)B} \log f(y_{(A)B}(h), p_{AB}(h)),\end{aligned} \tag{1}$$

where k_{AB} is the number of games that A against B at home; $\sum_{A(B)}$ denotes the sum over all games that A played against teams other than B in the league with A at home and $y_{A(B)}(h)$ the corresponding binary game outcomes; $\sum_{(A)B}$ is the sum over all games that B played against teams other than A when B is away and $y_{(A)B}(h)$ the corresponding outcomes. Let $\hat{p}_{AB}^{MWLE}(h)$ be the corresponding maximum weighted likelihood estimate (MWLE) of $p_{AB}(h)$. The MWLE of $p_{AB}(r)$ can be defined in a similar way.

We adopt the approximate Akaike criterion (Akaike, 1977, Akaike, 1985, and Hu and Zidek, 2002) to select the weights $\alpha_{AB}(h)$ and $\beta_{AB}(h)$ by minimizing with respect to both,

$$E(\hat{p}_{AB}(h) - p_{AB}(h))^2. \tag{2}$$

The resulting optima will, however, depend on the unknown p's being estimated. To address this problem we can use 'plug - in' estimators obtained in any reasonable way, for these p's, to obtain $\hat{\alpha}_{AB}(h)$ and $\hat{\beta}_{AB}(h)$ from Equation (2). One possible way of doing this is demonstrated in Section 3.

In most applications, we need confidence intervals (or the equivalent) for the parameters. The impossibility of finding exact confidence intervals based on the MWLE leads us to use approximate ones based on the asymptotic normality of the MWLE (see Theorem 5 of Hu, 1997). We obtain such a 95% confidence interval for $p_{AB}(h)$ as

$$[\hat{p}_{AB}^{MWLE}(h) - \hat{bias}_{AB} - 1.96\sqrt{\hat{var}_{AB}}, \hat{p}_{AB}^{MWLE}(h) + \hat{bias}_{AB} + 1.96\sqrt{\hat{var}_{AB}}]. \tag{3}$$

Here $\hat{bias}_{AB}$ and $\hat{var}_{AB}$ are the estimators of the bias and variance given in Theorem 5 of Hu (1997). With those estimates $\hat{p}_{AB}(h)$ and $\hat{p}_{BA}(r)$, we can find the predictive probabilities of winning, losing and drawing the game (along with their approximate confidence intervals) when a game is played at the home of Team A.

3 Predicting the NBA playoff results.

In this section, we turn to the problem of predicting the outcomes of NBA playoff games. Our analysis concerns the 1996-1997 season.

The home team advantage is significant in the NBA. We tested the null-hypothesis of no home team advantage against the alternative of a home team advantage and found a p-value of about 10^{-7} suggesting the need to separate home and away games.

Remark 1 *Here the simple sign test is used to test the home team advantage. There are* 29 *teams total in NBA. Each team played at home and away for* 41 *games. If a team has more wins at home, then the sign is* 1*. If a team has less wins at home, then the sign is* -1*. Use the data of 1996-97, there are* 28 *teams with more home wins and only one team with less home wins. By the sign test, the p-value is* $30 * 0.5^{-29} < 10^{-7}$.

To describe our application, let $Y_{AB}(h) \sim Bernoulli(1, p_{AB}(h))$ be independently distributed random variables representing a "win" or "loss" by team A in any one game played against team B while A is at home. We first estimate the predictive probabilities $p_{AB}(h)$ and $p_{BA}(h)$ where 'A' and 'B' denote respectively the Chicago Bulls and the Utah Jazz, two top NBA teams.

The use of the weighted likelihood seems especially appealing here given the paucity of "direct" information about the relative strengths of A and B. In fact, the Jazz played only one game in Chicago. The classical likelihood leaves no chance of finding reasonable parameter estimates. In contrast, the MWLE brings in information from games each of these teams played against others in the NBA. That is, the MWLE uses the information in the "relevant sample" in addition to that in the "direct sample".

We find the MWLE of $p_{AB}(h)$ (from the weighted likelihood (1)) to be

$$\hat{p}_{AB}^{MWLE}(h) = \bar{y}_{AB}(h) + \alpha_{AB}(h)(\bar{y}_{A(B)}(h) - \bar{y}_{AB}(h)) + \beta_{AB}(h)(\bar{y}_{(A)B}(h) - \bar{y}_{AB}(h)), \quad (4)$$

where $\bar{y}_{AB}(h)$ denotes the fraction of wins for A in the $k_{AB}(h)$ games played against B during the season with A at home. The $\bar{y}_{A(B)}(h)$ represents the corresponding fraction of wins for A in the $k_{A(B)}(h)$ games played against teams other than B with A at home.

By using the approximate Akaike criterion with a reasonable estimate $\hat{p}_{AB}(h)$ (described below), an optimal weight may be estimated by

$$\hat{\alpha}_{AB}(h) = \frac{V_{AB}(h)[V_{(A)B}(h) + (\bar{y}_{(A)B}(h) - \hat{p}_{AB}(h))(\bar{y}_{(A)B}(h) - \bar{y}_{A(B)}(h))]}{C + D} \tag{5}$$

and

$$\hat{\beta}_{AB}(h) = \frac{V_{AB}(h)[V_{A(B)}(h) + (\bar{y}_{A(B)}(h) - \hat{p}_{AB}(h))(\bar{y}_{A(B)}(h) - \bar{y}_{(A)B}(h))]}{C + D} \tag{6}$$

where

$$\begin{aligned}
V_{AB}(h) &= \frac{\hat{p}_{AB}(h)(1 - \hat{p}_{AB}(h))}{k_{AB}(h)}, \\
V_{A(B)}(h) &= \frac{\bar{y}_{A(B)}(h)(1 - \bar{y}_{A(B)}(h))}{k_{A(B)}(h)}, \\
V_{(A)B}(h) &= \frac{\bar{y}_{(A)B}(h)(1 - \bar{y}_{(A)B}(h))}{k_{(A)B}(h)}, \\
C &= V_{AB}(h)[V_{(A)B}(h) + V_{A(B)}(h) + (\bar{y}_{(A)B}(h) - \bar{y}_{A(B)}(h))^2]
\end{aligned}$$

and

$$D = V_{A(B)}(h)(\bar{y}_{(A)B}(h) - \hat{p}_{AB}(h))^2 + V_{(A)B}(h)(\bar{y}_{A(B)}(h) - \hat{p}_{AB}(h))^2 + V_{A(B)}(h)V_{(A)B}(h).$$

The corresponding mean square error of the MWLE may be estimated by

$$\begin{aligned}
M\hat{S}E_{MWLE} = {} & [\hat{\alpha}_{AB}(h)(\bar{y}_{A(B)}(h) - \hat{p}_{AB}^{MWLE}(h)) \\
& + \hat{\beta}_{AB}(h)(\bar{y}_{(A)B}(h) - \hat{p}_{AB}^{MWLE}(h))]^2 \\
& + \hat{\alpha}_{AB}^2(h)\frac{\bar{y}_{A(B)}(h)(1 - \bar{y}_{A(B)}(h))}{k_{A(B)}(h)} \\
& + \hat{\beta}_{AB}^2(h)\frac{\bar{y}_{(A)B}(h)(1 - \bar{y}_{(A)B}(h))}{k_{(A)B}(h)} \\
& + (1 - \hat{\alpha}_{AB}(h) - \hat{\beta}_{AB}(h))^2\frac{\hat{p}_{AB}^{MWLE}(h)(1 - \hat{p}_{AB}^{MWLE}(h))}{k_{AB}(h)}.
\end{aligned}$$

The 95% confidence interval of $p_{AB}(h)$ based on the MWLE would be:

$[\hat{p}_{AB}^{MWLE}(h) - \hat{bias}_{AB}(h) - 1.96\sqrt{\hat{var}_{AB}(h)}, \hat{p}_{AB}^{MWLE}(h) + \hat{bias}_{AB}(h) + 1.96\sqrt{\hat{var}_{AB}(h)}]$,

where

$$\hat{bias}_{AB}(h) = |\hat{\alpha}_{AB}(h)(\bar{y}_{A(B)}(h) - \hat{p}_{AB}^{MWLE}(h)) + \hat{\beta}_{AB}(h)(\bar{y}_{(A)B}(h) - \hat{p}_{AB}^{MWLE}(h))|$$

and

$$\begin{aligned} \hat{var}_{AB}(h) &= \hat{\alpha}_{AB}^2(h)\frac{\bar{y}_{A(B)}(h)(1-\bar{y}_{A(B)}(h))}{k_{A(B)}(h)} \\ &\quad + \hat{\beta}_{AB}^2(h)\frac{\bar{y}_{(A)B}(h)(1-\bar{y}_{(A)B}(h))}{k_{(A)B}(h)} \\ &\quad + (1-\hat{\alpha}_{AB}(h)-\hat{\beta}_{AB}(h))^2\frac{\hat{p}_{AB}^{MWLE}(h)(1-\hat{p}_{AB}^{MWLE}(h))}{k_{AB}(h)}. \end{aligned}$$

We now describe how we found the plug-in estimates, the optimal weights, the win probabilities and the corresponding confidence intervals by considering Bull against Jazz while Bull at home.

During the regular season, Bull played 41 games at home. One game was against Jazz and Bull win this game. So $k_{AB} = 1$ and $\bar{Y}_{AB} = 1$. Bull played 40 games against teams other than Jazz and win 38 of these games. Thus, $k_{A(B)} = 40$ and $\bar{Y}_{A(B)} = 0.95$. Jazz played 40 ($k_{(A)B} = 40$) games against teams other than Bull on road and win 26 of these games. $\bar{Y}_{(A)B} = 1 - 26/40 = 0.35$. For this case, the plug-in estimate,

$$\hat{p}_{AB}(h) = \frac{k_{AB}\bar{Y}_{AB} + k_{A(B)}\bar{Y}_{A(B)} + k_{(A)B}\bar{Y}_{(A)B}}{k_{AB} + k_{A(B)} + k_{(A)B}} = \frac{1+38+14}{1+40+40} = \frac{53}{81} = 0.6543.$$

The corresponding values in equation (5) and (6) can be calculated by using above results. And the values are: $V_{AB}(h) = 0.2262$, $V_{A(B)}(h) = 0,0011875$, $V_{(A)B}(h) = 0.0056875$, $C = 0,082987$ and $D = 0.000637$. Subsistute these values into equation (5) and (6), we get the optimal weights:

$$\hat{\alpha}_{AB}(h) = 0.50925, \text{ and } \hat{\beta}_{AB}(h) = 0.4831.$$

The MWLE in (1) is then

$$\hat{p}_{AB}^{MWLE}(h) = 0.66.$$

The corresponding mean square erroer, bias and variance of this MWLE are

$$MSE_{MWLE} = 0.001653,\ \hat{bias}_{AB}(h) = 0.002 \text{ and } \hat{var}_{AB}(h) = 0.001648.$$

The 95% confidence interval of $p_{AB}(h)$ based on this MWLE is then $[0.58, 0.74]$.

The above MWLE is based on the games with all teams that Bull played at home or Jazz played on the road. Each game has the same weight in the weighted likelihood. This seems unreasonable because some of the teams are significantly weaker than others. Now we only use the teams (10 teams in 1996/97 season) which won at least 50 games in the season. By using the games with these 10 teams, we calculate the win probabilities as well as the confidence intervals, which is deonted by MWLE1.

Before the 1996-97 finals between the Bulls and the Jazz, both teams had played the first and second round as well as the conference finals. This additional information is used in constructing MWLE2.

We now use MWLE, MWLE1 and MWLE2 to predict the 1996-97 Finals between the Bulls and Jazz. We report the point estimates of the probabilities, the mean square errors and the confidence intervals of $p_{AB}(h)$ in Table 1

Based on the probabilities and the confidence intervals of Table 1, we can find the probabilities with which the Bulls (and Jazz) will win the Final by Game 4, 5, 6 or 7. Also we can calculate the total win probabilities for the Bulls against the Jazz based on its home and away win probabilities given by each of the three estimation methods. Confidence intervals for these win probabilities may be obtained as well. In Table 2 where the results are reported, and in the tables that follow, that interval is obtained for any pair of teams say A and B from the 95% asymptotic intervals for A's home- and A's away-win-against-B probabilities. Since those intervals are stochastically dependent,

	MWLE	MWLE1	MWLE2
At Chicago	0.66 (0.002)	0.77 (0.007)	0.75 (0.004)
95% C.I.	[0.58,0.74]	[0.60,0.94]	[0.62,0.89]
At Utah	0.40 (0.002)	0.36 (0.008)	0.34 (0.004)
95% C.I.	[0.32,0.48]	[0.16,0.55]	[0.21,0.47]

Table 1: The Bulls predictive win probabilities (with mean square error) and confidence intervals based on MWLE, MWLE1 and MWLE2 for a future game between the Bulls and the Jazz during the 1996 - 97 Season.

we use a Bonferonni argument and obtain an asymptotic interval of confidence at least 90%. In obtaining that interval, we rely on the heuristically obvious fact that the overall win probability must be a monotonically increasing function of the home and away win probabilities.

Table 1 indicates general agreement between MWLE1 and MWLE2. But MWLE gives a much smaller estimator of a Chicago win at home. Both MWLE1 and MWLE2 predict that the Bulls would win the Finals with high probability. Also MWLE1 and MWLE2 predict the Bulls will win at Game 6. These predictions agree with the actual result: the Bulls won the Final at Game 6.

To explore the performance of our method further, we calculated prediction probabilities for other pairs of teams, the Bulls vs. the Miami Heat, the Atlanta Hawks, the New York Knicks as well as the Miami Heat against the Knicks. The results are summarized in Table 3.

From Table 4, we see that MWLE, MWLE1 and MWLE2, respectively predict a Bulls win against the Heat with probabilities 0.62, 0.72 and 0.77. MWLE1 and MWLE2 also predict that most probably the Bulls will win at Game 5. That prediction proved to be correct.

	Game #	Game 4	Game 5	Game 6	Game 7	Total	90+%[a] C.I.
MWLE	Bull Win	0.07	0.11	0.21	0.21	0.61	[0.43,0.77]
	Jazz Win	0.04	0.13	0.11	0.11	0.39	[0.23,0.56]
MWLE1	Bull Win	0.07	0.11	0.27	0.26	0.71	[0.30,0.95]
	Jazz Win	0.02	0.11	0.08	0.08	0.29	[0.05,0.70]
MWLE2	Bull Win	0.07	0.10	0.26	0.26	0.69	[0.37,0.92]
	Jazz Win	0.02	0.12	0.09	0.08	0.31	[0.08,0.63]

Table 2: The predictive probabilities of a Bulls win against the Jazz together with confidence intervals for MWLE, MWLE1 and MWLE2 in the 1996 - 97 Final.

[a]An interval of confidence at least 90% based on an asymptotic approximation.

Table 5 gives MWLE, MWLE1 and MWLE2 predictions when the Bulls play the Atlanta Hawks. MWLE1 and MWLE2 predict that the Bulls will win with probabilities 0.96 and 0.93 respectively while MWLE is only 0.72. MWLE1 and MWLE2 also predict a Bulls win at game 5 with the highest probabilities (0.43 and 0.40). [In the playoffs the Bulls did win at game 5.]

Table 7 shows that a Heat - Knicks game will be close. MWLE and MWLE1 predict that the Heat has a slight advantage in the playoffs, while MWLE2 favors the Knicks slightly. In fact, the Heat won at game 7. However an accident occurred in that series leading to a suspension of several New York players in Games 6 and 7. Undoubtedly this influenced the outcome.

Overall, MWLE is more conservative in that it's predictions are closer to 0.5 than the other methods. This is because MWLE uses some not-so-relevant information from games involving weak teams. When the Bulls and the Jazz play weak teams each wins. Thus, these data will tend to increase both of their success rates. However, since they both enjoy that benefit, the relevant difference in their estimated strengths will

Pairs		MWLE	MWLE1	MWLE2
Bull	At Chicago	0.59 (0.001)	0.72 (0.007)	0.75 (0.005)
vs	95% C.I.	[0.52,0.67]	[0.55,0.90]	[0.60,0.90]
Heat	At Miami	0.51 (0.002)	0.44 (0.01)	0.45 (0.007)
	95% C.I.	[0.41,0.61]	[0.24,0.64]	[0.29,0.61]
Bull	At Chicago	0.73 (0.002)	0.95 (0.001)	0.91 (0.003)
vs	95% C.I.	[0.65,0.82]	[0.88,1.0]	[0.81,1.0]
Hawks	At Atlanta	0.32 (0.002)	0.32 (0.009)	0.44 (0.007)
	95% C.I.	[0.34,0.51]	[0.23,0.62]	[0.28,60]
Bull	At Chicago	0.66 (0.002)	0.76 (0.006)	0.77 (0.004)
vs	95% C.I.	[0.58,0.74]	[0.59,0.93]	[0.63,90]
Knicks	At New York	0.49 (0.002)	0.42 (0.009)	0.46 (0.006)
	95% C.I.	[0.39,0.58]	[0.23,0.62]	[0.29,0.60]
Heat	At Miami	0.52 (0.003)	0.63 (0.012)	0.59 (0.008)
vs	95% C.I.	[0.40,0.63]	[0.38,0.87]	[0.33,0.85]
Knicks	At New York	0.51 (0.002)	0.40 (0.013)	0.34 (0.008)
	95% C.I.	[0.42,0.60]	[0.20,0.60]	[0.16,0.53]

Table 3: The predictive win probabilities, the mean square errors and the confidence intervals based on MWLE, MWLE1 and MWLE2: the Bulls against the Miami Heat (where the Bulls win); the Bulls against the Atlanta Hawks (the Bulls win); the Bulls against the New York Knicks (the Bulls win) and the Miami Heat against the New York Knicks (the Heat wins).

	Game #	Game 4	Game 5	Game 6	Game 7	Total	90+%[a] C.I.
MWLE	Bull Win	0.09	0.18	0.17	0.18	0.62	[0.44,0.79]
	Heat Win	0.04	0.08	0.13	0.13	0.38	[0.21,0.56]
MWLE1	Bull Win	0.10	0.24	0.16	0.22	0.72	[0.32,0.97]
	Heat Win	0.02	0.05	0.12	0.08	0.28	[0.03,0.68]
MWLE2	Bull Win	0.11	0.27	0.17	0.22	0.77	[0.43,0.96]
	Heat Win	0.02	0.03	0.11	0.07	0.23	[0.04,0.57]

Table 4: The predictive win probabilities and the confidence intervals for MWLE, MWLE1 and MWLE2 between the Bulls and Miami Heat for the 1996-97 East Conference Final.

[a]An interval of confidence at least 90% based on an asymptotic approximation.

	Game #	Game 4	Game 5	Game 6	Game 7	Total	90+%[a] C.I.
MWLE	Bull Win	0.09	0.24	0.16	0.23	0.72	[0.54,0.87]
	Hawks Win	0.02	0.05	0.13	0.08	0.28	[0.13,0.46]
MWLE1	Bull Win	0.16	0.43	0.15	0.21	0.96	[0.79,1.00]
	Hawks Win	0.00	0.00	0.03	0.01	0.04	[0.00,0.21]
MWLE2	Bull Win	0.16	0.40	0.16	0.21	0.93	[0.72,1.00]
	Hawks Win	0.00	0.00	0.05	0.02	0.07	[0.00,0.28]

Table 5: The predictive probability of a win and the confidence intervals for MWLE, MWLE1 and MWLE2 between the Bulls and the Atlanta Hawks during the 1996-97 playoffs.

[a]An interval of confidence at least 90% based on an asymptotic approximation.

	Game #	Game 4	Game 5	Game 6	Game 7	Total	90+%[a] C.I.
MWLE	Bull Win	0.10	0.22	0.17	0.20	0.69	[0.50,0.85]
	Knicks Win	0.03	0.06	0.12	0.10	0.31	[0.15,0.50]
MWLE1	Bull Win	0.10	0.26	0.16	0.23	0.75	[0.35,0.98]
	Knicks Win	0.02	0.04	0.12	0.07	0.25	[0.02,0.65]
MWLE2	Bull Win	0.12	0.28	0.16	0.22	0.78	[0.46,0.96]
	Knicks Win	0.01	0.03	0.11	0.07	0.22	[0.04,0.54]

Table 6: The predictive probability of a win along with the confidence intervals of MWLE, MWLE1 and MWLE2 between the Bulls and the New York Knicks during the 1996-97 playoffs.

[a]An interval of confidence at least 90% based on an asymptotic approximation.

	Game #	Game 4	Game 5	Game 6	Game 7	Total	90+%[a] C.I.
MWLE	Heat Win	0.07	0.14	0.16	0.16	0.53	[0.31,0.74]
	Knicks Win	0.06	0.11	0.15	0.15	0.47	[0.26,0.69]
MWLE1	Heat Win	0.06	0.16	0.14	0.20	0.56	[0.12,0.92]
	Knicks Win	0.05	0.09	0.18	0.12	0.44	[0.08,0.88]
MWLE2	Heat Win	0.04	0.13	0.11	0.19	0.47	[0.07,0.91]
	Knicks Win	0.07	0.12	0.21	0,13	0.53	[0.09,0.93]

Table 7: The predictive probability of win and the confidence intervals of MWLE, MWLE1 and MWLE2 between the Miami Heat and the New York Knicks during the 1996-97 playoff.

[a]An interval of confidence at least 90% based on an asymptotic approximation.

diminish, making the MWLE tend toward 0.5. MWLE1 and MWLE2 agree with each other, the latter giving slightly more precise predictions (as measured by the length of the associated predictive intervals in Table 1) because it incorporates the playoff games.

The Bulls and the Knicks did not meet in the playoffs. However, MWLE1 and MWLE2 predict a hypothetical Bulls win with probabilities 0.75 and 0.78 had they met. Both predict a hypothetical Bulls win for the series at game 5.

4 Concluding Remarks.

The method in this paper provides guidelines for the development of a prediction strategy. Its implementation, more specifically the construction of weights entails the incorporation of any special features that may obtain when the game is played. For example, one might need to incorporate the knowledge that certain key players cannot play in that game. [This last consideration did arise in the playoff between the Miami Heat and the New York Knicks.]

The need for the incorporation of such features was reaffirmed by an unpublished analysis carried out in the summer of 1998 by Farouk Nathoo. In that analysis, he twice simulated the entire 1997/98 season based on the previous year's results. In his report he compared the simulation results with the actual results. Among other things he found the fraction of wins for each of the 29 NBA teams and for example we include the results for the Atlantic Division and give these results in Table 8.

We see in this example that the simulated winning precentages are in reasonable agreement with the actual results except in the case of the Nets, the Knicks and the Celtics. Given the severity of the challenge of predicting the outcomes of all games over an entire year, we find our results encouraging.

Team	Win % : Actual	Win % Simulation 1	Win % : Simulation 2
Heat	67	66	59
Nets	52	35	38
Knicks	52	65	66
Wizards	51	52	54
Magic	50	54	48
Celtics	44	20	26
Sixers	38	27	35

Table 8: The percentage of wins in the actual and two simulated 1997/98 season for the NBA's Atlantic Division based on the WL win probability estimators obtained at the end of the previous season.

The WL method can be applied in other sports such as baseball, hockey (see Hu and Zidek 2002), soccer. In this paper, we chose the same weight for all teams. This seems unreasonable in some cases and there we may be able to use the rank of the teams to get better weights. This is another topic for the future.

Finally, we would note the abundance of alternative approaches, Bayesian (Berger, 1985) and non-Bayesian that could be used in this context. Some specific methods were discribed in the Introduction, We intend to compare our approach with some of these in future work. Here, we restricted our comparisons to an extension of one of the non - Bayesian approaches based on that of Bradley and Terry (1952) to estimate the probabilities of a Bulls win for both home and away games against the Jazz. (We found the corresponding probabilities for the remaining teams as well but do not report them here.) With these probabilities we could then compute the termination probabilities analogous to those in Table 2.

To be more precise, we fitted a logistic model using the software R with the response variable being 1 or 0 according as the outcome of any game during the season was a visitor or home victory. We used dummy variables to represent visitor and home teams in each game throughout the season. Thus for example, BullsV=1 and Supersonics=1, all other dummies being 0, would mean those two teams were playing for that particular game, the visitors being the Bulls. For each of the factors, 'visitor' and 'home' we represented by the dummies in this way, we arbitrarily chose the 76er's as the baseline team. Thus, in effect, the fitted intercept, suitably transformed, provides an estimate of the likelihood of a '1' in the purely hypothetical situation where the 76er's played themselves at home as the visitors. The coefficients for the remaining dummies represent the deviations from the 76er's performance for each of the other teams depending on whether they were playing at home or away. In summary, the R model we fitted is given by:

```
glm(formula = outcome ~ BucksV + BucksH + BulletsV + BulletsH +
    BullsV + BullsH + CavaliersV + CavaliersH + CelticsV + CelticsH +
    ClippersV + ClippersH + GrizzliesV + GrizzliesH + HawksV +
    HawksH + HeatV + HeatH + HornetsV + HornetsH + JazzV + JazzH +
    KingsV + KingsH + KnicksV + KnicksH + LakersV + LakersH +
    MagicV + MagicH + MavericksV + MavericksH + NetsV + NetsH +
    NuggetsV + NuggetsH + PacersV + PacersH + PistonsV + PistonsH +
    RaptorsV + RaptorsH + RocketsV + RocketsH + SpursV + SpursH +
    SunsV + SunsH + SupersonicsV + SupersonicsH + TimberwolvesV +
    TimberwolvesH + TrailBlazersV + TrailBlazersH + WarriorsV +
    WarriorsH, family = binomial, data = nba)
```

The results differed somewhat from those obtained by the MWLE2 WL method. To be specific we found the probability of a Bulls win at home to be 0.76 as compared with

	Game #	Game 4	Game 5	Game 6	Game 7
MWLE2	Bull Win	0.07	0.10	0.26	0.26
	Jazz Win	0.02	0.12	0.09	0.08
Bradley-Terry	Bull Win	0.05	0.08	0.25	0.28
	Jazz Win	0.03	0.14	0.09	0.09

Table 9: The predictive probabilities of a Bulls win against the Jazz for both the MWLE2 and Bradley-Terry (logistic) based methods in the 1996 - 97 Final.

the 0.75 seen in Table 1 while the corresponding probabilities for the Jazz were 0.71 and 0.66 respectively. These differences became more pronounced when we computed the probabilities corresponding to Table 2. We see a comparison of the results in Table 9.

In Table 9, we see that the Bradley - Terry extension points to a Bulls victory on Game 7 while the MWLE2 is ambivalent between games 6 and 7. Obviously a more extensive comparison would be needed to assess the relative performance of the methods. But considering the large number of parameters needed by the logistic model, these very preliminary results make the weighted likelihood model more desirable for forecasting the outcomes of NBA playoff games.

Although in this manuscript we have used only binary outcome information about team wins or losses, the theory can be extended to incorporate more complex outcome information such as the scores, for example. In that case, we could have defined the $Y_{AB}(h)$ to be the score of team A against team B when team A is at home and so on.

Acknowledgements. We are indebted to Farouk Nathoo for comments that improved the clarity of our presentation.

References

[1] Akaiki, H. (1977). On entropy maximization principle. In P. R. Krishnaiah, (ed.), *Applications of Statistics.* Amsterdam: North-Holland, 27-41.

[2] Akaiki, H. (1985). Prediction and entropy. *A Celebration of statistics, The ISI Centenary Volume*. Berlin, Spring-Verlag.

[3] Bennett, J (Ed) (1998). *Statistics in sport.*, London, Arnold Applications of Statistics.

[4] Berger, J.O. (1985). *Statistical decision theory and Bayesian analysis, Second edition.* New York: Springer - Verlag.

[5] Bradley, R. A. and Terry, M. E. (1952). Rank analysis of incomplete block designs - I. The method of paired comparisons. *Biometrika*, 39, 324 - 345.

[6] Carlin, BP (1996). Improved NCAA basketball tournament modelling via point spread and team strength information. *The Amer Statistician*, 50, 39-43..

[7] Harville, DA (1977). The use of linear-model methodology to rate high school or college football teams. *JASA*, 72, 278-289.

[8] Harville, DA (1980). Predictions for national football league games via linear-model methodology. *JASA*, 75, 516-524.

[9] Hu, F. (1994). Relevance weighted smoothing and a new bootstrap method. Ph.D. Thesis, Department of Statistics, University of British Columbia.

[10] Hu, F. (1997). Asymptotic properties of relevance weighted likelihood estimations. *Canadian Journal of Statistics*, 25, 45-60.

[11] Hu, F. and Rosenberger, W.F. (2000). Anaylysis of time trends in adaptive designs with application to a neurophysiology experiment. *Statistics in Medicine*, 19, 2067-2075.

[12] Hu, F and Zidek J V (1993b). Relevant samples and their information. *Unpublished manuscript*.

[13] Hu, F. and Zidek, J.V. (2001). The relevance weighted likelihood with applications. In *Empirical Bayes and Likelihood Inference.* S.E.Ahmed and N.M.Reid, eds.) New York: Springer, 211-234.

[14] Hu, F. and Zidek, J.V. (2002). The weighted likelihood. *Canadian Journal of Statistics*, 30, 347-371.

[15] Hu, F, Rosenberger, W.F. and Zidek, JV (2000). The relevance weighted likelihood for dependent data. *Metrika.* 51, 223-43.

[16] Kuonen, D and Roehrl, ASA (2000). Was France's World Cup win pure chance?. *Student*. 3, 153-166.

[17] Schwertman, NC, Schenk, L, and Holbrook, BC (1993). More Probability models for the NCAA regional baketball tournaments. Technical Report, Department of Mathematics and Statistics, California State University, Chico.

[18] Schwertman, NC, Schenk, L, and Holbrook, BC (1996). More probability models for the NCAA regional baketball tournaments. *The Amer Statistician*, 50, 34-38.

[19] Stern, H (1992). Who's number one? - rating football teams. In *Proceedings of the Section on Sports Statistics (Vol 1).* Alexandria, VA: American Statistical Association, pp 1-6.

[20] Van Eeden, C, and Zidek, JV (2000). Combining the data from two normal populations to estimate the mean of one when their means diffence is bounded. Available from http://hajek.stat.ubc.ca/ jim/pubs. Submitted.

[21] Van Eeden, C, and Zidek, JV (2002). Combining sample information in estimating ordered normal means. *Sankyha, Series A*, 64, 588-610.

[22] Wang, S. (2001). Maximum relevance weighted likelihood estimation. Ph. D. Thesis, Department of Statistics, University of British Columbia.