General

Late-Game Reversals in Professional Basketball, Football, and Hockey

Paramjit S. GILL

Empirical evidence suggests that in basketball, football, and hockey, the leader at the beginning of the final period (quarter or period) wins the game about 80% of the time. We discuss modeling of late-game reversals in NBA, NFL, and NHL sports. The models are built around the assumptions that basketball scores and football scores are normally distributed and hockey scores vary according to a Poisson distribution. The models also accommodate the proverbial home field advantage. We use data from the 1997–1998 regular seasons of the leagues to estimate the parameters for the models. Predictions from the probabilistic models are in excellent agreement with the actual outcomes.

KEY WORDS: Goodness-of-fit; Normal distribution; Poisson distribution; Sports data.

1. INTRODUCTION

Most of us enjoy watching our favorite teams playing (and winning) in professional sports. Suppose you are watching on TV your favorite team playing a late night game. If your team is leading in the late part of the game, can you afford to switch the TV off to go to sleep? Mosteller (1997) suggested that one can do so unless the game is very close. Empirical evidence suggests that in basketball, football, and hockey, the leader at the beginning of the final period (quarter) wins the game about 80% of the time (Cooper, DeNeve, and Mosteller 1992). A late-game reversal occurs if the team trailing at the beginning of the final period recovers to win the game.

In this article we revisit the question of reversals using some recent and more extensive data from games played during the regular seasons of the National Basketball Association (NBA), the National Football League (NFL), and the National Hockey League (NHL). An NBA game consists of four quarters of 12 minutes each. If the scores at the end of regular time are tied, play is continued in five-minute overtime periods until the tie is broken. Similarly, an NFL game is played in four quarters of 15 minutes each, and a tied game is extended for a sudden-death period of 15 minutes. The game can end in a draw. Ice hockey in the NHL is played in three periods of 20 minutes each. In case of tied scores at the end of regular time, the game goes for a sudden-death tie breaker period of five minutes. About one in six of regular season NHL games ends in a tie.

The probabilistic modeling is built around the assumptions that basketball scores and football scores are normally distributed and hockey scores vary according to a Poisson distribution. The use of normal distribution for the American football (NFL and college) and basketball data has a long history (Stern 1991, 1998; Carlin 1996). Poisson distribution is an appropriate model for low scoring sports like ice hockey and soccer. Mullet (1977) suggested using Poisson modeling for NHL scores. Recently, Danehy and Lock (1995) developed the College Hockey Offensive/Defensive Ratings (CHODR) model based on Poisson distribution. Dixon and Coles (1997) and Lee (1997) used the Poisson model for the soccer scores in English leagues.

Almost all the league games are played at the "home arena" of one of the two teams. Thus, the two teams in a game are distinguished as "home" and "away" teams. It is well known that professional games exhibit some home-field advantage. We do not distinguish between the favorites and underdogs. As noted by Cooper, DeNeve, and Mosteller (1992), the present data also exhibit home-field advantage in reversals. Team scores at the beginning of the last period

Paramjit S. Gill is a Professor, Department of Mathematics and Statistics, Okanagan University College, Kelowna, BC, Canada V1V 1V7 (E-mail: pgill@okanagan.bc.ca). He also has an adjunct appointment as a research associate at the University of British Columbia, Vancouver. This work was supported by an NSERC grant. The results reported in this article were presented at the 1998 Joint Statistical Meetings in Dallas, Texas. The author thanks Tim Swartz of Simon Fraser University, two referees, and an associate editor for their comments and suggestions that improved the presentation of the material.

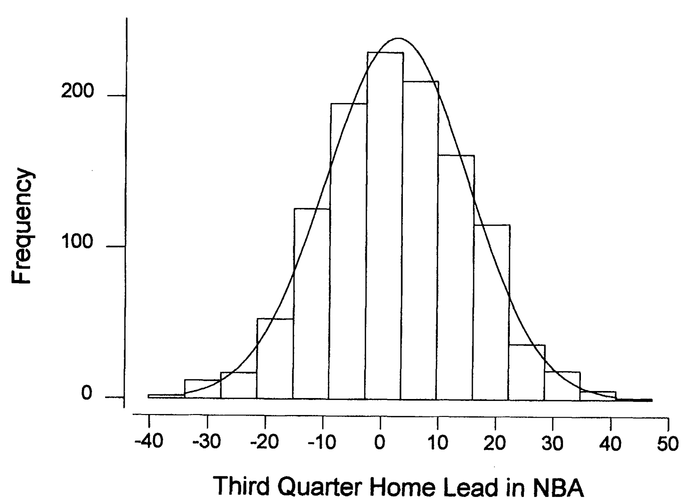

Figure 1. Normal Probability Fit to the Distribution of Third Quarter Home Lead in NBA Games.

and the scores during the last period are modeled separately. As two referees pointed out, various teams in a given league differ in their ability to score and to produce late-game reversals. However, for the sake of simplicity and lack of sufficient data, we do not accommodate the differential ability of the teams.

We use data from the 1997–1998 regular seasons of the NBA, NFL, and NHL. This covered 1065 NHL games, 1,188 NBA games, and 240 NFL games. The predictions from the probabilistic models are compared with the actual outcomes. The data were gathered from the Web site http://www.sportingnews.com/

2. BASKETBALL

Let X and Y denote, respectively, the home team and the away team scores at the end of the third quarter. We assume a bivariate normal distribution for (X, Y) so that the difference $Z = X - Y$ is also modeled by a normal distribution, say $N(\mu, \sigma)$. Note that we are assuming a common distribution for all the teams regardless of their unequal ability. For the 1997–1998 NBA data, the normal fit is good (Figure 1, Anderson-Darling normality test $p = .07$). For these data, $\mu = 2.6$ and $\sigma = 12.3$.

As the scores are discrete data, a continuity correction is needed for the use of a normal distribution. A score of n points is taken to be equivalent to the interval $(n-.5, n+.5)$ on the continuous scale. With this continuity correction, the probability of a tie at the end of three quarters is given by

$$\begin{aligned}\Pr(\text{Tie}) &= \Pr(-.5 < Z < .5) \\ &= \Phi\{(.5-\mu)/\sigma\} - \Phi\{(-.5-\mu)/\sigma\} = .032,\end{aligned} \tag{1}$$

where $\Phi(.)$ denotes the cumulative distribution function for the standard normal probability distribution. The actual proportion of games tied at the end of three quarters was .036 (43 of the 1,188 games). As we are interested in the fourth-quarter reversals, we ignore the games that were tied at the end of three quarters. Among the untied games, let $\Pr(H_n)$ denote the probability that the home team leads by a difference of n (> 0) points at the end of three quarters, and let $\Pr(A_n)$ denote the corresponding probability for the away team. Then

$$\begin{aligned}\Pr(H_n) &= \frac{\Pr(n-.5 < Z < n+.5)}{1-\Pr(\text{Tie})} \\ &= \frac{\Phi\{(n+.5-\mu)/\sigma\} - \Phi\{(n-.5-\mu)/\sigma\}}{1-\Phi\{(.5-\mu)/\sigma\} + \Phi\{(-.5-\mu)/\sigma\}},\end{aligned} \tag{2}$$

Table 1. Fourth Quarter Home Lead in NBA Games

	Status at three quarters: Home leads	Status at three quarters: Away leads	Overall
Mean	−.27	1.19	.34
St. Dev	7.15	7.32	7.28

and

$$\begin{aligned}\Pr(A_n) &= \frac{\Pr(-n-.5 < Z < -n+.5)}{1-\Pr(\text{Tie})} \\ &= \frac{\Phi\{(-n+.5-\mu)/\sigma\} - \Phi\{(-n-.5-\mu)/\sigma\}}{1-\Phi\{(.5-\mu)/\sigma\} + \Phi\{(-.5-\mu)/\sigma\}}.\end{aligned} \tag{3}$$

The probability that the home team leads at the end of three quarters is then $\Pr(H) = \sum_{n=1}^{\infty} \Pr(H_n) = .586$ as compared to the actual proportion of .576 (660 home leads in 1,145 games). The probability that the away team leads at the end of three quarters is $\Pr(A) = \sum_{n=1}^{\infty} \Pr(A_n) = .414$ as compared to the actual proportion of .424 (485 out of 1,145 games).

At the end of the fourth quarter, three possible outcomes are: scores are tied, home team wins, or the away team wins. Let U and V denote, respectively, the home team and away team scores during the fourth quarter and the overtime, if any. Under the assumption of bivariate normality for (U, V), the difference $W = U - V$ is also modeled by a normal distribution. It is interesting to observe that the performances of the two teams in the last quarter are very similar. The mean difference was only .34 points in favor of the home team. But if we look at the fourth quarter difference conditional on the game status at the end of three quarters, some interesting features emerge (Table 1). If the home team had a lead at three quarters, their lead decreased, on average, by .27 points. And if the visitors were leading at three quarters, the home team decreased the lead by an average of 1.19 points. This suggests that teams trailing at the beginning of the fourth quarter played better to come back. In doing so the home team demonstrated home court advantage. Accordingly, we have the following two situations:

Case 1. When the home team led at three quarters, let $N(\delta_1, \nu_1)$ denote the distribution for difference W. Obviously, the chances of keeping the lead in the fourth quarter depend on the third quarter lead. Given that the home team had a lead of n (> 0) points at three quarters, the conditional probability that the home team wins the game is given by

$$\begin{aligned}&\Pr(\text{Home team wins the game} \mid H_n) \\ &\qquad = \Pr(W > -n) = 1 - \Phi\{(-n-\delta_1)/\nu_1\}\end{aligned} \tag{4}$$

Case 2. When the away team led at three quarters, let $N(\delta_2, \nu_2)$ denote the distribution of the difference W. In this case

$$\begin{aligned}&\Pr(\text{Away team wins the game} \mid A_n) \\ &\qquad = \Pr(W < n) = \Phi\{(n-\delta_2)/\nu_2\}.\end{aligned} \tag{5}$$

The probabilities shown in (4)–(5) are of interest to the fans watching the game. It should be emphasized that there are not enough cases to estimate these probabilities by proportions. The use of a probability model is therefore a must. Figure 2 shows these probabilities graphically for various values of n. To compute these probabilities, we used the parameter values $\delta_1 = -.27$, $\nu_1 = 7.15$, $\delta_2 = 1.19$, $\nu_2 = 7.32$, as given in Table 1.

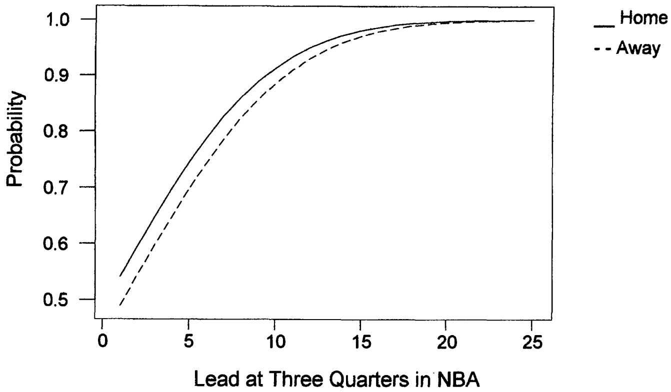

Figure 2. Probability that the Third Quarter Leader Wins an NBA Game.

The probability that home team keeps the lead is then

$$\Pr(\text{Home team wins the game}|\text{Home team led at three quarters}) = \frac{\sum_{n=1}^{\infty}[1-\Phi\{(-n-\delta_1)/\nu_1\}]\Pr(H_n)}{\sum_{n=1}^{\infty}\Pr(H_n)} = .852 \quad (6)$$

Note that the numerator in (6) is the probability of intersection of two events: {home team leads at three quarters} $\bigcap$ {home team wins the game}. The denominator is the probability that home team led at three quarters. The probability of reversal by the away team is then $1 - .852 = .148$.

The probability that away team keeps the lead is

$$\Pr(\text{Away team wins the game}|\text{Away team led at three quarters}) = \frac{\sum_{n=1}^{\infty}\Phi\{(n-\delta_2)/\nu_2\}\Pr(A_n)}{\sum_{n=1}^{\infty}\Pr(A_n)} = .788 \quad (7)$$

And the probability of reversal by the home team is $1 - .788 = .212$. Using (6) and (7), we get the probability that the leader at the third quarter wins the game as $(.586)(.852) + (.414)(.788) = .825$ as compared to the actual proportion of .837 (959 out of 1,145 games).

Table 2. Goodness-of-fit for Outcomes in NBA Games

	Status at three quarters			
	Home leads		*Away leads*	
Final outcome	*Observed*	*Expected*	*Observed*	*Expected*
Home wins	578	563	104	103
Away wins	82	97	381	382

Table 2 shows the expected and observed number of outcomes in 1,145 games in the 1997–1998 NBA regular season. Looking at the actual outcomes, we find that the home team is about twice as likely to stage a fourth-quarter reversals than is the away team (21.4% versus 12.4%). The corresponding percentages reported by Cooper et al. (1992), based on a sample of 189 NBA games from the 1990–1991 regular season, were 33.3% and 10.5%. We also note that the observed number of late-game reversals by the home team are in excellent agreement with the prediction, but the visitors were not as successful in reversals as predicted by our model. One possible reason for this discrepancy could be the enormous home crowd support in the games when home team was leading by a narrow margin. If data were available, it would be interesting to investigate reversals in the last five minutes of the game.

3. AMERICAN FOOTBALL

In NFL games, the team scores in a quarter are typically skewed to the right. But the differences (= *home score* − *away score*) have relatively symmetric distribution. Unlike Stern (1991), we do not distinguish between the favorite and underdog teams. Let X and Y denote, respectively, the home team and the away team scores at the end of the third quarter. We assume that the difference $Z = X - Y$ is modeled by a normal distribution, say $N(\mu, \sigma)$. A normal probability fitting for our data shows that the assumption of normality for Z is reasonable (Figure 3, Anderson-Darling normality test $p = .32$). The parameter values are estimated as $\mu = 2.1$ and $\sigma = 13.1$. Of the 240 games in the 1997–1998 NFL season, 13 were tied at three quarters and another 3 ended in a draw. These 16 games are excluded from the further analysis.

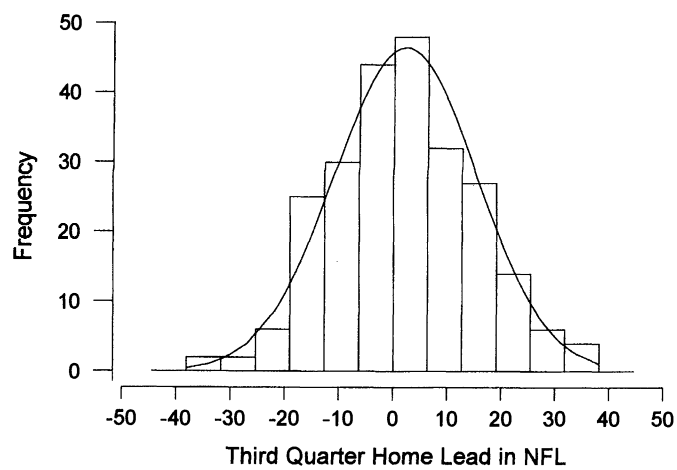

Figure 3. Normal Probability Fit to the Distribution of Third Quarter Home Lead in NFL Games.

Let U and V denote, respectively, the home team and away team scores during the fourth quarter and the overtime, if any. Under the assumption of bivariate normality for (U, V), the difference $W = U - V$ is also modeled by a normal distribution. As in the case of basketball, the parameter values for the distribution of W are assumed dependent on the status of the third quarter game (Table 3).

Table 3. Fourth Quarter Home Lead in NFL Games

	Status at three quarters		
	Home leads	*Away leads*	*Overall*
Mean	−.4	1.4	.5
St. Dev	7.2	7.1	7.2

When the home team was leading at three quarters, the away decreased the margin in the fourth quarter, on average, by .4 points. And when the away team was leading at three quarters, the home team decreased the margin in the fourth quarter by 1.4 points. We use the same arguments as in case of NBA games. The probability that the home team wins the game given that they were leading at three quarters is computed to be .864. The corresponding probability for the visitors is .801. The overall probability that the leader at the third quarter wins the game is .835 as compared to the actual proportion of .821 (184 out of 224).

Table 4. Goodness-of-fit for Outcomes in NFL games

	Status at three quarters			
	Home leads		Away leads	
Final outcome	Observed	Expected	Observed	Expected
Home wins	112	111	23	19
Away wins	17	18	72	76

Table 4 shows the number of fourth quarter reversals in the 1997–1998 NFL season. The observed number of reversals by the visitors is very close to the prediction but our model predicted a lower number of home team comebacks than were observed. We see that home team was twice as likely as visiting team to make fourth quarter reversals (24.2% versus 13.2%). In comparison, based on a sample of 93 NFL games from the 1990–1991 regular season, Cooper et al. (1992) reported that both teams were almost equally likely to stage a fourth quarter comeback.

4. ICE HOCKEY

A game in NHL consists of three periods of 20 minutes each. The number of goals scored by each game during a fixed time interval can be modeled using independent Poisson processes (Mullet, 1977). During the course of a game, each team makes, on average, about 25 shots at the goal with about 10% success rate. Therefore, the event of a goal score can be considered "rare" and the Poisson assumption is reasonable.

Let X and Y denote, respectively, the home team and the away team scores at the end of the second period. We assume that X has a Poisson distribution with mean λ and Y has a Poisson distribution with mean μ independent of X. Therefore, the joint distribution of (X, Y) is

$$\Pr(X = x, Y = y) = \frac{e^{-\lambda}\lambda^x}{x!} \times \frac{e^{-\mu}\mu^y}{y!}; \quad x = 0, 1, 2, \ldots; \quad y = 0, 1, 2, \ldots \quad (8)$$

Table 5 shows the observed joint and marginal distributions for X and Y for our data. The expected marginal distributions were computed using $\lambda = 1.770$ and $\mu = 1.694$ as estimated from the same data. Chi-square goodness-of-fit test does not reject the hypothesis of independence of Poisson distributions for X and Y ($\chi^2 = 25.2$ at 24 degrees of freedom with $p = .40$). Poisson distributions fit excellently for X ($\chi^2 = 2.2$ at 4 df with $p = .70$) and Y ($\chi^2 = 1.3$ at 4 df with $p = .86$).

Using the Poisson joint probability model (8), we can find the probabilities of the three possible outcomes at the end of the second period. The probability of a tie is $\Pr(\text{Tie}) = \sum_{k=0}^{\infty} \Pr(X = k, Y = k)$. The probability that the home team leads by n (> 0) goals is $\Pr(H_n) = \sum_{k=0}^{\infty} \Pr(X = k + n, Y = k)$ and the probability that the away team leads by n (> 0) goals is given by $\Pr(A_n) = \sum_{k=0}^{\infty} \Pr(X = k, Y = k + n)$. Table 6 gives the computed probabilities, the expected and observed number of games for various outcomes in two periods. The agreement between the number of observed and expected frequencies is excellent ($\chi^2 = 5.6$ at 9 df with $p = .78$).

Table 6. Status at the End of Second Period in NHL Games

		Number of Games	
Outcome	Probability	Expected	Observed
Tie	.224	239	224
Home Lead			
1 Goal	.192	204	208
2 Goals	.121	129	145
3 Goals	.059	63	64
$\geq$ 4 Goals	.032	34	25
Away Lead			
1 Goal	.184	196	199
2 Goals	.110	117	118
3 Goals	.051	54	55
$\geq$ 4 Goals	.027	29	27
Total	1.000	1065	1065

At the end of the third period, three possible outcomes are: scores are tied, home team wins, or the away team

Table 5. Distributions of Second Period Scores in NHL Games

Away team		Home team score (X)					Marginal for Y	
score (Y)		0	1	2	3	≥4	Observed	Expected
0		34	68	45	24	16	187	195
1		58	90	88	69	42	347	331
2		46	85	71	42	30	274	281
3		27	49	42	26	15	159	159
≥4		10	40	22	18	8	98	98
Marginal	Observed	175	332	268	179	111		
for X	Expected	181	321	284	168	111		

wins. About 75% of the games going overtime end in a tie. In the following, the term "third period" will mean the regular time third period plus overtime, if any. Let U and V denote, respectively, the home team and away team scores in the third period. Again, we assume that U and V are independent Poisson random variables. However, the mean parameters for the Poisson distributions are assumed different depending on the game status at the end of the second period. This incorporates the home-away team differential in late-game reversals. Table 7 shows the third period mean scores of home and away teams broken down according to the status at the second period. Thus, we observe that a team leading at the second period performed, on average, better than the other team in the third period also. Home team shows an advantage over the away team in keeping the lead and in game reversals.

Table 7. Third Period Mean Scores in NHL Games

	Status at two periods		
Team	*Tie*	*Home leads*	*Away leads*
Home	.969	.989	.900
Away	.875	.765	.910
No. of Games	224	442	399

Table 8. Probability that Second Period Leader Wins an NHL Game

	Leader: home team		Leader: away team	
Lead	*Leader wins*	*Reversal*	*Leader wins*	*Reversal*
1	.727	.082	.667	.117
2	.916	.019	.883	.031
3	.981	.003	.969	.006
4	.997	.000	.994	.001

Given that the home team lead by n (> 0) goals at the second period, the probabilities for the three possible outcomes at the end of the game are:

$$\Pr(\text{Home team wins } |H_n) = \sum_{k=0}^{\infty} \Pr(U > k-n, V = k),$$

$$\Pr(\text{Away team wins } |H_n) = \sum_{k=0}^{\infty} \Pr(U < k-n, V = k),$$

and

$$\Pr(\text{Tie } |H_n) = \sum_{k=0}^{\infty} \Pr(U = k-n, V = k).$$

Similarly, we can compute probabilities for all possible combinations of scenario at the second period and at the end of the game. Table 8 shows the computed conditional probabilities of the leader victory and reversal for $n = 1, 2, 3, 4$.

Now, the probability that the home team keeps the lead is given by

$$\Pr(\text{Home team wins} \mid \text{Home team led at two periods}) = \frac{\sum_{n=1}^{\infty} \Pr(\text{Home team wins } |H_n) \Pr(H_n)}{\sum_{n=1}^{\infty} P(H_n)} = .839$$

The probability of reversal by the away team is

$$\Pr(\text{Away team wins} \mid \text{Home team led at two periods}) = \frac{\sum_{n=1}^{\infty} \Pr(\text{Away team wins } |H_n) \Pr(H_n)}{\sum_{n=1}^{\infty} \Pr(H_n)} = .047$$

And the probability that the away team ties the scores is $1 - .839 - .047 = .114$. Similarly, the probability that the away team keeps the lead is computed to be equal to .796, and the probability of reversal by the home team is .068. Based on these conditional probabilities, the expected number of outcomes for various scenarios can be computed. Table 9 shows the expected and observed frequencies for various outcomes for the 1997–1998 NHL season. The number of outcomes are reasonably close to those predicted by the Poisson probability model. Late-game reversals in hockey are not as dramatic as in case of basketball and football. There seem to be two obvious reasons for this—slow pace of scoring and possibility of a game ending in a tie. Counting ties also among the reversals (each team gets one point from a tie), home teams staged about 21% reversals and away teams managed about 15% reversals. In this regard, our model gives an excellent prediction.

Table 9. Goodness-of-fit for Outcomes in NHL Games

	Status at the second period					
	Home leads		Away leads		Tie	
Final outcome	*Observed*	*Expected*	*Observed*	*Expected*	*Observed*	*Expected*
Home wins	374	371	23	27	84	81
Away wins	23	21	315	318	70	70
Tie	45	50	61	54	70	72

[Received March 1998. Revised July 1999.]

REFERENCES

Carlin, B. P. (1996), "Improved NCAA Basketball Tournament Modeling via Point Spread and Team Strength Information," *The American Statistician*, 50, 39–43.

Cooper, H., DeNeve, K. M., and Mosteller F. (1992), "Predicting Professional Sports Game Outcomes from Intermediate Game Scores," *Chance*, 5, 18–22.

Danehy, T., and Lock, R. (1995), "CHODR: Using Statistics to Predict College Hockey," *STATS*, 13, 10–14.

Dixon, M. J., and Coles, S. G. (1997), "Modelling Association Football Scores and Inefficiencies in the Football Betting Market," *Applied Statistics*, 46, 165–280.

Lee, A. J. (1997), "Modeling Scores in the Premier League: Is Manchester United Really the Best?," *Chance*, 10, 15–19.

Mosteller, F. (1997), "Lessons from Sports Statistics," *The American Statistician*, 51, 305–310.

Mullet, G. M. (1977), "Simeon Poisson and the National Hockey League," *The American Statistician*, 31, 8–12.

Stern, H. (1991), "On the Probability of Winning a Football Game," *The American Statistician*, 45, 179–183.

——— (1998), "A Statistician Reads the Sports Pages," *Chance*, 11, 17–21.