A Brownian Motion Model for the Progress of Sports Scores

Hal S. STERN*

The difference between the home and visiting teams' scores in a sports contest is modeled as a Brownian motion process defined on $t \in (0, 1)$, with drift μ points in favor of the home team and variance σ^2. The model obtains a simple relationship between the home team's lead (or deficit) ℓ at time t and the probability of victory for the home team. The model provides a good fit to the results of 493 professional basketball games from the 1991–1992 National Basketball Association (NBA) season. The model is applied to the progress of baseball scores, a process that would appear to be too discrete to be adequately modeled by the Brownian motion process. Surprisingly, the Brownian motion model matches previous calculations for baseball reasonably well.

KEY WORDS: Baseball; Basketball; Probit regression

1. INTRODUCTION

Sports fans are accustomed to hearing that "team A rarely loses if ahead at halftime" or that "team B had just accomplished a miracle comeback." These statements are rarely supported with quantitative data. In fact the first of the two statements is not terribly surprising; it is easy to argue that approximately 75% of games are won by the team that leads at halftime. Suppose that the outcome of a half-game is symmetrically distributed around 0 so that each team is equally likely to "win" the half-game (i.e., assume that two evenly matched teams are playing). In addition, suppose that the outcomes of the two halves of a game are independent and identically distributed. With probability .5 the same team will win both half-games, and in that case the team ahead at halftime certainly wins the game. Of the remaining probability, it seems plausible that the first half winner will defeat the second half winner roughly half the time. This elementary argument suggests that in contests among fairly even teams, the team ahead at halftime should win roughly 75% of the time. Evaluating claims of "miraculous" comebacks is more difficult. Cooper, DeNeve, and Mosteller (1992) estimated the probability that the team ahead after three quarters of the game eventually wins the contest for each of the four major sports (basketball, baseball, football, hockey). They found that the leading team won more than 90% of the time in baseball and about 80% of the time in the other sports. They also found that the probability of holding a lead is different for home and visiting teams. Neither the Cooper, et al. result nor the halftime result described here considers the size of the lead, an important factor in determining the probability of a win.

The goal here is to estimate the probability that the home team in a sports contest wins the game given that they lead by ℓ points after a fraction $t \in (0, 1)$ of the contest has been completed. Of course, the probability for the visiting team is just the complement. The main focus is the game of basketball.

Among the major sports, basketball has scores that can most reasonably be approximated by a continuous distribution. A formula relating ℓ and t to the probability of winning allows for more accurate assessment of the propriety of certain strategies or substitutions. For example, should a star player rest at the start of the fourth quarter when his team trails by 8 points or is the probability of victory from this position too low to risk such a move? In Section 2 a Brownian motion model for the progress of a basketball score is proposed, thereby obtaining a formula for the probability of winning conditional on ℓ and t. The model is applied to the results of 493 professional basketball games in Section 3. In Section 4 the result is extended to situations in which it is known only that $\ell > 0$. Finally, in Section 5 the Brownian motion model is applied to a data set consisting of the results of 962 baseball games. Despite the discrete nature of baseball scores and baseball "time" (measured in innings), the Brownian motion model produces results quite similar to those of Lindsey (1977).

2. THE BROWNIAN MOTION MODEL

To begin, we transform the time scale of all sports contests to the unit interval. A time $t \in (0, 1)$ refers to the point in a sports contest at which a fraction t of the contest has been completed. Let $X(t)$ represent the lead of the home team at time t. The process $X(t)$ measures the difference between the home team's score and the visiting team's score at time t; this may be positive, negative, or 0. Westfall (1990) proposed a graphical display of $X(t)$ as a means of representing the results of a basketball game. Naturally, in most sports (including the sport of most interest here, basketball), $X(t)$ is integer valued. To develop the model, we ignore this fact, although we return to it shortly. We assume that $X(t)$ can be modeled as a Brownian motion process with drift μ per unit time ($\mu > 0$ indicates a μ point per game advantage for the home team) and variance σ^2 per unit time. Under the Brownian motion model,

$$X(t) \sim N(\mu t, \sigma^2 t)$$

and $X(s) - X(t)$, $s > t$, is independent of $X(t)$ with

$$X(s) - X(t) \sim N(\mu(s - t), \sigma^2(s - t)).$$

* Hal S. Stern is Associate Professor, Department of Statistics, Harvard University, Cambridge, MA 02138. Partial support for this work was provided by Don Rubin's National Science Foundation Grant SES-8805433. The author thanks Tom Cover for suggesting the problem and the halftime argument several years ago and Elisabeth Burdick for the baseball data. Helpful comments were received from Tom Belin, Andrew Gelman, and Carl Morris.

Journal of the American Statistical Association
September 1994, Vol. 89, No. 427, Statistics in Sports

The probability that the home team wins a game is $\Pr(X(1) > 0) = \Phi(\mu/\sigma)$, and thus the ratio μ/σ indicates the magnitude of the home field advantage. In most sports, the home team wins approximately 55–65% of the games, corresponding to values of μ/σ in the range .12–.39. The drift parameter μ measures the home field advantage in points (typically thought to be 3 points in football and 5–6 points in basketball).

Under the random walk model, the probability that the home team wins [i.e., $X(1) > 0$] given that they have an ℓ point advantage (or deficit) at time t [i.e., $X(t) = \ell$] is

$$\begin{aligned} P_{\mu,\sigma}(\ell, t) &= \Pr(X(1) > 0 \mid X(t) = \ell) \\ &= \Pr(X(1) - X(t) > -\ell) \\ &= \Phi\left(\frac{\ell + (1-t)\mu}{\sqrt{(1-t)\sigma^2}}\right), \end{aligned}$$

where Φ is the cdf of the standard normal distribution. Of course, as $t \to 1$ for fixed $\ell \neq 0$, the probability tends to either 0 or 1, indicating that any lead is critically important very late in a game. For fixed t, the lead ℓ must be relatively large compared to the remaining variability in the contest for the probability of winning to be substantial.

The preceding calculation treats $X(t)$ as a continuous random variable, although it is in fact discrete. A continuity correction is obtained by assuming that the observed score difference is the value of $X(t)$ rounded to the nearest integer. If we further assume that contests tied at $t = 1$ [i.e., $X(1) = 0$] are decided in favor of the home team with probability .5, then it turns out that

$$P^{cc}_{\mu,\sigma}(\ell, t) = 0.5\,\Phi\left(\frac{\ell - .5 + (1-t)\mu}{\sqrt{(1-t)\sigma^2}}\right) + .5\,\Phi\left(\frac{\ell + .5 + (1-t)\mu}{\sqrt{(1-t)\sigma^2}}\right).$$

In practice, the continuity correction seems to offer little improvement in the fit of the model and causes only minor changes in the estimates of μ and σ. It is possible to obtain a more accurate continuity correction that accounts for the drift in favor of the home team in deciding tied contests. In this case .5 is replaced by a function of μ, σ, and the length of the overtime used to decide the contest.

The Brownian motion model motivates a relatively simple formula for $P_{\mu,\sigma}(\ell, t)$, the probability of winning given the lead ℓ and elapsed time t. A limitation of this formula is that it does not take into account several potentially important factors. First, the probability that a home team wins, conditional on an ℓ point lead at time t, is assumed to be the same for any basketball team against any opponent. Of course, this is not true; Chicago (the best professional basketball team during the period for which data has been collected here) has a fairly good chance of making up a 5-point halftime deficit ($\ell = -5$, $t = .50$) against Sacramento (one of the worst teams), whereas Sacramento would have much less chance of coming from behind against Chicago. One method for taking account of team identities would be to replace μ with an estimate of the difference in ability between the two teams in a game, perhaps the Las Vegas point spread. A second factor not accounted for is whether the home team is in possession of the ball at time t and thus has the next opportunity to score. This is crucial information in the last few minutes of a game ($t > .96$ in a 48-minute basketball game). Despite the omission of these factors, the formula appears to be quite useful in general, as demonstrated in the remainder of the article.

3. APPLICATION TO PROFESSIONAL BASKETBALL

Data from professional basketball games in the United States are used to estimate the model parameters and to assess the fit of the formula for $P_{\mu,\sigma}(\ell, t)$. The results of 493 National Basketball Association (NBA) games from January to April 1992 were obtained from the newspaper. This sample size represents the total number of games available during the period of data collection and represents roughly 45% of the complete schedule. We assume that these games are representative of modern NBA basketball games (the mean score and variance of the scores were lower years ago). The differences between the home team's score and the visiting team's score at the end of each quarter are recorded as $X(.25)$, $X(.50)$, $X(.75)$, and $X(1.00)$ for each game. For the ith game in the sample, we also represent these values as $X_{i,j}$, $j = 1, \ldots, 4$. The fourth and final measurement, $X(1.00) = X_{i,4}$, is the eventual outcome of the game, possibly after one or more overtime periods have been played to resolve a tie score at the end of four quarters. The overtime periods are counted as part of the fourth quarter for purposes of defining X. This should not be a problem, because $X(1.00)$ is not used in obtaining estimates of the model parameters. In a typical game, on January 24, 1992, Portland, playing Atlanta at home, led by 6 points after one quarter and by 9 points after two quarters, trailed by 1 point after three quarters, and won the game by 8 points. Thus $X_{i,1} = 6$, $X_{i,2} = 9$, $X_{i,3} = -1$, and $X_{i,4} = 8$.

Are the data consistent with the Brownian motion model? Table 1 gives the mean and standard deviation for the results of each quarter and for the final outcome. In Table 1 the outcome of quarter j refers to the difference $X_{i,j} - X_{i,j-1}$ and the final outcome refers to $X_{i,4} = X(1.00)$. The first three quarters are remarkably similar; the home team outscores the visiting team by approximately 1.5 points per quarter, and the standard deviation is approximately 7.5 points. The fourth quarter seems to be different; there is only a slight advantage to the home team. This may be explained by the fact that if a team has a comfortable lead, then it is apt to ease up or use less skillful players. The data suggests that the home team is much more likely to have a large lead after three quarters; this may explain the fourth quarter results in Table 1. The normal distribution appears to be a satisfactory approximation to the distribution of score differences in each quarter, as indicated by the QQ plots in Figure 1. The cor-

Table 1. Results by Quarter of 493 NBA Games

Quarter	Variable	Mean	Standard deviation
1	$X(.25)$	1.41	7.58
2	$X(.50) - X(.25)$	1.57	7.40
3	$X(.75) - X(.50)$	1.51	7.30
4	$X(1.00) - X(.75)$	.22	6.99
Total	$X(1.00)$	4.63	13.18

relations between the results of different quarters are negative and reasonably small ($r_{12} = -.13$, $r_{13} = -.04$, $r_{14} = -.01$, $r_{23} = -.06$, $r_{24} = -.05$, and $r_{34} = -.11$). The standard error for each correlation is approximately .045, suggesting that only the correlation between the two quarters in each half of the game, r_{12} and r_{34}, are significantly different from 0. The fact that teams with large leads tend to ease up may explain these negative correlations, a single successful quarter may be sufficient to create a large lead. The correlation of each individual quarter's result with the final outcome is approximately .45. Although the fourth quarter results provide some reason to doubt the Brownian motion model, it seems that the model may be adequate for the present purposes. We proceed to examine the fit of the formula $P_{\mu,\sigma}(\ell, t)$ derived under the model.

The formula $P_{\mu,\sigma}(\ell, t)$ can be interpreted as a probit regression model relating the game outcome to the transformed variables $\ell/\sqrt{1-t}$ and $\sqrt{1-t})$ with coefficients $1/\sigma$ and μ/σ. Let $Y_i = 1$ if the home team wins the ith game [i.e., $X(1) > 0$)] and 0 otherwise. For now, we assume that the three observations generated for each game, corresponding to the first, second, and third quarters, are independent. Next we investigate the effect of this independence assumption. The probit regression likelihood L can be expressed as

$$L = \prod_{i=1}^{493} \prod_{j=1}^{3} \Phi\left(\frac{X_{ij} + \left(1 - \frac{j}{4}\right)\mu}{\sqrt{\left(1 - \frac{j}{4}\right)\sigma^2}} \right)^{Y_i} \times \left(1 - \Phi\left(\frac{X_{ij} + \left(1 - \frac{j}{4}\right)\mu}{\sqrt{\left(1 - \frac{j}{4}\right)\sigma^2}} \right) \right)^{(1-Y_i)}$$

$$= \prod_{i=1}^{493} \prod_{j=1}^{3} \Phi\left(\alpha \frac{X_{ij}}{\sqrt{1 - \frac{j}{4}}} + \beta \sqrt{1 - \frac{j}{4}} \right)^{Y_i} \times \left(1 - \Phi\left(\alpha \frac{X_{ij}}{\sqrt{1 - \frac{j}{4}}} + \beta \sqrt{1 - \frac{j}{4}} \right) \right)^{(1-Y_i)}$$

where $\alpha = 1/\sigma$ and $\beta = \mu/\sigma$. Maximum likelihood estimates of α and β (and hence μ and σ) are obtained using a Fortran program to carry out a Newton–Raphson procedure. Convergence is quite fast (six iterations), with $\hat{\alpha} = .0632$ and $\hat{\beta} = .3077$ implying

$$\hat{\mu} = 4.87 \quad \text{and} \quad \hat{\sigma} = 15.82.$$

An alternative method for estimating the model parameters directly from the Brownian motion model, rather than through the implied probit regression, is discussed later in this section. Approximate standard errors of $\hat{\mu}$ and $\hat{\sigma}$ are obtained via the delta method from the asymptotic variance and covariance of $\hat{\alpha}$ and $\hat{\beta}$:

$$\text{s.e.}(\hat{\mu}) = .90 \quad \text{and} \quad \text{s.e.}(\hat{\sigma}) = .89.$$

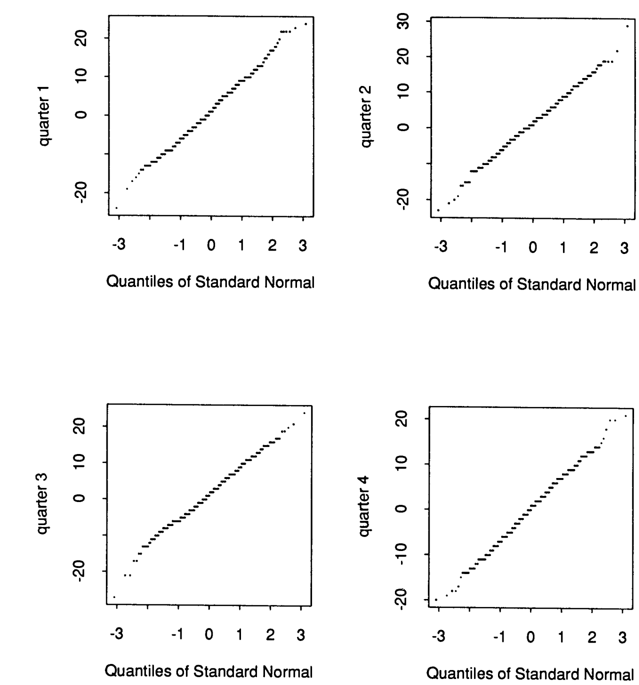

Figure 1. Q-Q Plots of Professional Basketball Score Differences by Quarter. These are consistent with the normality assumption of the Brownian motion model.

These standard errors are probably somewhat optimistic, because they are obtained under the assumption that individual quarters contribute independently to the likelihood, ignoring the fact that groups of three quarters come from the same game and have the same outcome Y_i. We investigate the effect of the independence assumption by simulation using two different types of data. "Nonindependent" data, which resemble the NBA data, are obtained by simulating 500 Brownian motion basketball games with fixed μ, σ and then using the three observations from each game (the first, second, and third quarter results) to produce data sets consisting of 1,500 observations. Independent data sets consisting of 1,500 independent observations are obtained by simulating 1,500 Brownian motion basketball games with fixed μ, σ and using only one randomly chosen quarter from each game. Simulation results using "nonindependent" data suggest that parameter estimates are approximately unbiased but the standard errors are 30–50% higher than under the independence condition. The standard errors above are computed under the assumption of independence and are therefore too low. Repeated simulations, using "nonindependent" data with parameters equal to the maximum likelihood estimates, yield improved standard error estimates, s.e.$(\hat{\mu}) = 1.3$ and s.e.$(\hat{\sigma}) = 1.2$.

The adequacy of the probit regression fit can be measured relative to the saturated model that fits each of the 158 different (ℓ, t) pairs occurring in the sample with its empirical probability. Twice the difference between the log-likelihoods is 134.07, which indicates an adequate fit when compared to the asymptotic chi-squared reference distribution with 156 degrees of freedom. As is usually the case, there is little difference between the probit regression results and logistic regression results using the same predictor variables. We use probit regression to retain the easy interpretation of the regression coefficients in terms of μ, σ. The principal contribution of the Brownian motion model is that regressions based on the transformations of (ℓ, t) suggested by the Brownian motion model, $(\ell/\sqrt{1-t}, \sqrt{1-t})$, appear to provide a better fit than models based on the untransformed variables. As mentioned in Section 2, it is possible to fit the Brownian motion model with a continuity correction. In this case the estimates for μ and σ are 4.87 and 15.80, almost identical to the previous estimates. For simplicity, we do not use the continuity correction in the remainder of the article.

Under the Brownian motion model, it is possible to obtain estimates of μ, σ without performing the probit regression. The game statistics in Table 1 provide direct estimates of the mean and standard deviation of the assumed Brownian process. The mean estimate, 4.63, and the standard deviation estimate, 13.18, obtained from Table 1 are somewhat smaller than the estimates obtained by the probit model. The differences can be attributed in part to the failure of the Brownian motion model to account for the results of the fourth quarter. The probit model appears to produce estimates that are more appropriate for explaining the feature of the games in which we are most interested—the probability of winning.

Table 2 gives the probability of winning for several values of ℓ, t. Due to the home court advantage, the home team has a better than 50% chance of winning even if it is behind by two points at halftime ($t = .50$). Under the Brownian motion model, it is not possible to obtain a tie at $t = 1$ so this cell is blank; we might think of the value there as being approximately .50. In professional basketball $t = .9$ corresponds roughly to 5 minutes remaining in the game. Notice that home team comebacks from 5 points in the final 5 minutes are not terribly unusual. Figure 2 shows the probability of winning given a particular lead; three curves are plotted corresponding to $t = .25, .50, .75$. In each case the empirical probabilities are displayed as circles with error bars (± two binomial standard errors). To obtain reasonably large sample sizes for the empirical estimates, the data were divided into bins containing approximately the same number of games (the number varies from 34 to 59). Each circle is plotted at the median lead of the observations in the bin. The model appears consistent with the pattern in the observed data.

Figure 3 shows the probability of winning as a function of time for a fixed lead ℓ. The shape of the curves is as expected. Leads become more indicative of the final outcome as time passes and, of course, larger leads appear above smaller leads. The $\ell = 0$ line is above .5, due to the drift in favor of the home team. A symmetric graph about the horizontal line at .5 is obtained if we fix $\mu = 0$. Although the probit regression finds μ is significantly different than 0, the no drift model $P_{0,\sigma}(\ell, t) = \Phi(\ell/\sqrt{(1-t)\sigma^2})$ also provides a reasonable fit to the data with estimated standard deviation 15.18.

Figure 4 is a contour plot of the function $P_{\hat{\mu},\hat{\sigma}}(\ell, t)$ with time on the horizontal axis and lead on the vertical axis. Lines on the contour plot indicate game situations with equal probability of the home team winning. As long as the game is close, the home team has a 50–75% chance of winning.

4. CONDITIONING ONLY ON THE SIGN OF THE LEAD

Informal discussion of this subject, including the introduction to this article, often concerns the probability of winning given only that a team is ahead at time t ($\ell > 0$) with the exact value of the lead unspecified. This type of partial information may be all that is available in some circumstances. Integrating $P_{\mu,\sigma}(\ell, t)$ over the distribution of the lead ℓ at time t yields (after some transformation)

Table 2. $P_{\hat{\mu},\hat{\sigma}}(\ell, t)$ for Basketball Data

	Lead						
Time t	$\ell = -10$	$\ell = -5$	$\ell = -2$	$\ell = 0$	$\ell = 2$	$\ell = 5$	$\ell = 10$
.00				.62			
.25	.32	.46	.55	.61	.66	.74	.84
.50	.25	.41	.52	.59	.65	.75	.87
.75	.13	.32	.46	.56	.66	.78	.92
.90	.03	.18	.38	.54	.69	.86	.98
1.00	.00	.00	.00		1.00	1.00	1.00

$$P_{\mu/\sigma}(t) = \Pr(X(1) > 0 \mid X(t) > 0)$$

$$= \int_0^\infty \Phi\left(y\sqrt{\frac{t}{1-t}} + \sqrt{1-t}\,\frac{\mu}{\sigma}\right)\frac{1}{\sqrt{2\pi}} \times \exp\left(-\left(y - \sqrt{t}\,\frac{\mu}{\sigma}\right)^2\Big/2\right) dy,$$

which depends only on the parameters μ and σ through the ratio μ/σ. The integral is evaluated at the maximum likelihood estimates of μ and σ using a Fortran program to implement Simpson's rule. The probability that the home team wins given that it is ahead at $t = .25$ is .762, the probability at $t = .50$ is .823, and the probability at $t = .75$ is .881. The corresponding empirical values, obtained by considering only those games in which the home team led at the appropriate time point, (263 games for $t = .25$, 296 games for $t = .50$, 301 games for $t = .75$) are .783, .811, and .874, each within a single standard error of the model predictions.

If it is assumed that $\mu = 0$, then we obtain the simplification

$$P_0(t) = \frac{1}{2} + \frac{1}{\pi}\tan^{-1}\left(\sqrt{\frac{t}{1-t}}\right),$$

with $P_0(.25) = 2/3$, $P_0(.50) = 3/4$, and $P_0(.75) = 5/6$. Because there is no home advantage when $\mu = 0$ is assumed, we combine home and visiting teams together to obtain empirical results. We find that the empirical probabilities (based on 471, 473, and 476 games) respectively are .667, .748, and .821. Once again, the empirical results are in close agreement with the results from the probit model.

5. OTHER SPORTS

Of the major sports, basketball is best suited to the Brownian motion model because of the nearly continuous nature of the game and the score. In this section we report the results of applying the Brownian motion model to the results of the 1986 National League baseball season. In baseball, the teams play nine innings; each inning consists of two half-innings, with each team on offense in one of the half-innings. The half-inning thus represents one team's opportunity to score. The average score for one team in a single half-inning is approximately .5. More than 70% of the half-innings produce 0 runs. The data consist of 962 games (some of the National League games were removed due to data entry errors or because fewer than nine innings were played).

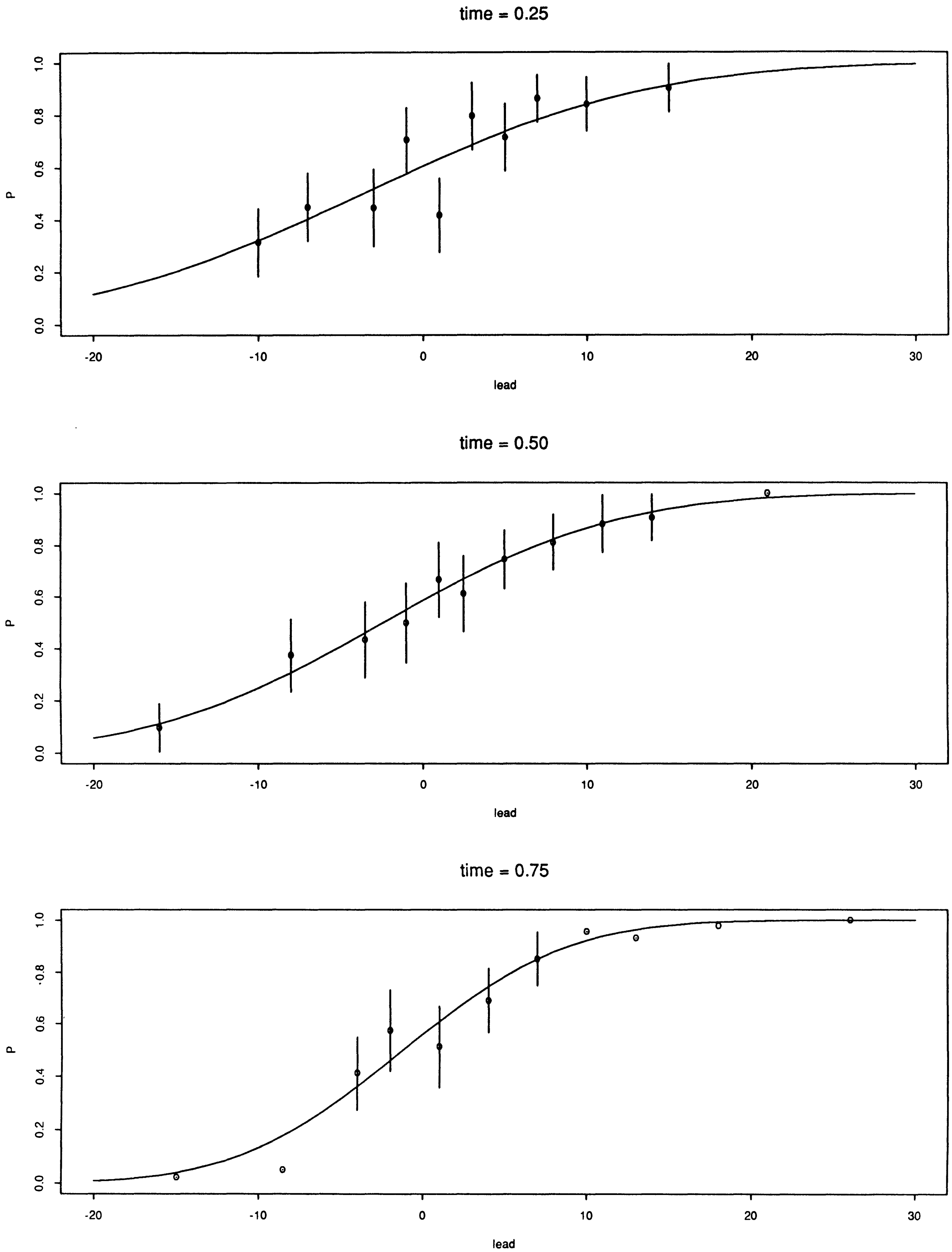

Figure 2. Smooth Curves Showing Estimates of the Probability of Winning a Professional Basketball Game, $P_{\hat{\mu},\hat{\sigma}}(\ell, t)$, as a Function of the Lead ℓ under the Brownian Motion Model. The top plot is $t = .25$, the middle plot is $t = .50$; and the bottom plot is $t = .75$. Circles $\pm$ two binomial standard errors are plotted indicating the empirical probability. The horizontal coordinate of each circle is the median of the leads for the games included in the calculations for the circle.

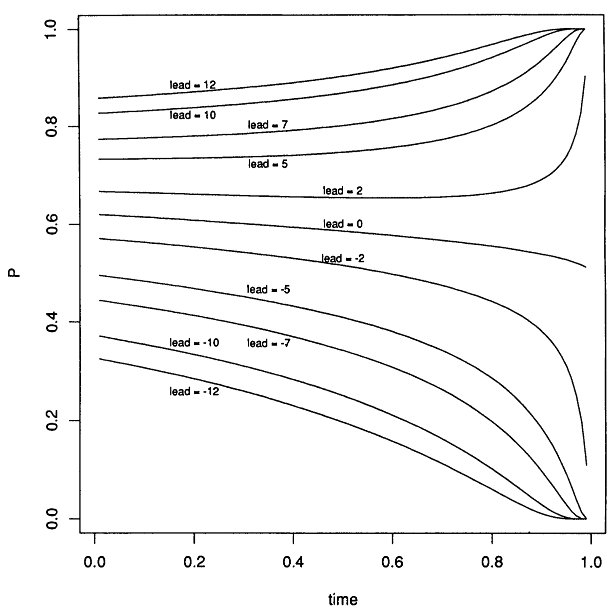

Figure 3. Estimated Probability of Winning a Professional Basketball Game, $P_{\hat{\mu},\hat{\sigma}}(\ell, t)$, as a Function of Time t for Leads of Different Sizes.

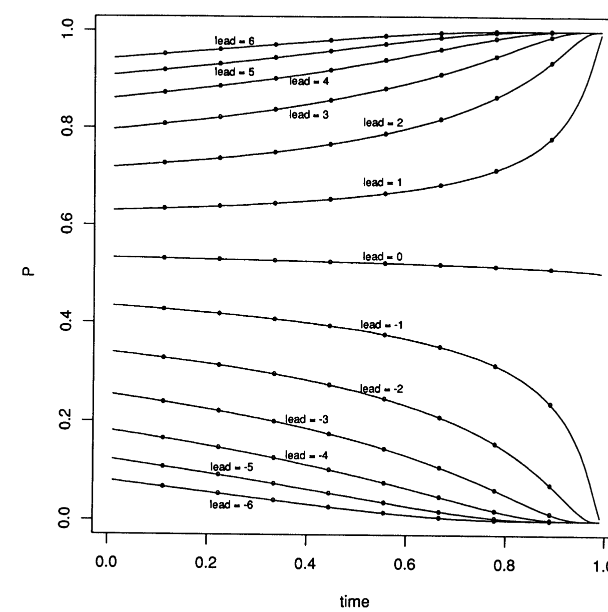

Figure 5. Estimated Probability of Winning a Baseball Game, $P_{\hat{\mu},\hat{\sigma}}(\ell, t)$, as a Function of Time t for Leads of Different Sizes.

Clearly, the Brownian motion model is not tailored to baseball as an application, although one might still consider whether it yields realistic predictions of the probability of winning given the lead and the inning. Lindsey (1961, 1963, 1977) reported a number of summary statistics, not repeated here, concerning the distribution of runs in each inning. The innings do not appear to be identically distributed due to the variation in the ability of the players who tend to bat in a particular inning. Nevertheless, we fit the Brownian motion model to estimate the probability that the home team wins given a lead ℓ at time t (here $t \in \{1/9, \ldots, 8/9\}$). The probit regression obtains the point estimates $\hat{\mu} = .34$ and $\hat{\sigma} = 4.04$. This mean and standard deviation are in good agreement with the mean and standard deviation of the margin of victory for the home team in the data. The asymptotic standard errors for $\hat{\mu}$ and $\hat{\sigma}$ obtained via the delta method are .09 and .10. As in the basketball example, these standard errors are optimistic, because each game is assumed to contribute eight independent observations to the probit regression likelihood, when the eight observations from a single game share the same outcome. Simulations suggest that the standard error of $\hat{\mu}$ is approximately .21 and the standard error of $\hat{\sigma}$ is approximately .18. The likelihood ratio test statistic, comparing the probit model likelihood to the saturated model, is 123.7 with 170 degrees of freedom. The continuity correction again has only a small effect.

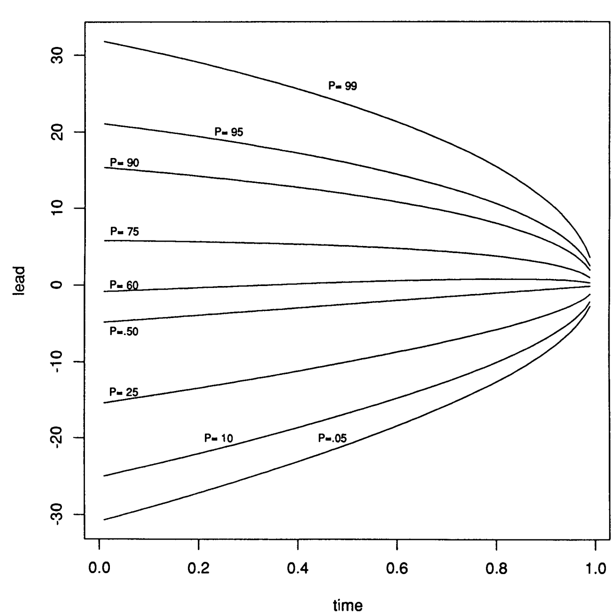

Figure 4. Contour Plot Showing Combinations of Home Team Lead and Fraction of the Game Completed for Which the Probability of the Home Team Winning is Constant for Professional Basketball Data.

Figure 5 shows the probability of winning in baseball as a function of time for leads of different sizes; circles are plotted at the time points corresponding to the end of each inning,

Table 3. $P_{\hat{\mu},\hat{\sigma}}(\ell, t)$ for Baseball Compared to Lindsey's Results

t	ℓ	$\hat{\mu} = .34$, $\hat{\sigma} = 4.04$	$\hat{\mu} = .0$, $\hat{\sigma} = 4.02$	Lindsey
3/9	0	.53	.50	.50
3/9	1	.65	.62	.63
3/9	2	.75	.73	.74
3/9	3	.84	.82	.83
3/9	4	.90	.89	.89
5/9	0	.52	.50	.50
5/9	1	.67	.65	.67
5/9	2	.79	.77	.79
5/9	3	.88	.87	.88
5/9	4	.94	.93	.93
7/9	0	.52	.50	.50
7/9	1	.71	.70	.76
7/9	2	.86	.85	.88
7/9	3	.95	.94	.94
7/9	4	.98	.98	.97

$t \in \{1/9, \ldots, 8/9\}$. Despite the continuous curves in Figure 5, it is not possible to speak of the probability that the home team wins at times other than those indicated by the circles, because of the discrete nature of baseball time. We can compare the Brownian motion model results with those of Lindsey (1977). Lindsey's calculations were based on a Markov model of baseball with transition probabilities estimated from a large pool of data collected during the late 1950s. He essentially assumed that $\mu = 0$. Table 3 gives a sample of Lindsey's results along with the probabilities obtained under the Brownian motion model with $\mu = 0$ ($\hat{\sigma} = 4.02$ in this case) and the probabilities obtained under the Brownian motion model with μ unconstrained. The agreement is fairly good. The inadequacy of the Brownian motion model is most apparent in late game situations with small leads. The Brownian motion model does not address the difficulty of scoring runs in baseball, because it assumes that scores are continuous. Surprisingly, the continuity correction does not help. We should note that any possible model failure is confounded with changes in the nature of baseball scores between the late 1950s (when Lindsey's data were collected) and today. The results in Table 3 are somewhat encouraging for more widespread use of the Brownian motion model.

[Received May 1993. Revised July 1993.]

REFERENCES

Cooper, H., DeNeve, K. M., and Mosteller, F. (1992), "Predicting Professional Game Outcomes From Intermediate Game Scores," *CHANCE*, 5, 3-4, 18-22.

Lindsey, G. R. (1961), "The Progress of the Score During a Baseball Game," *Journal of the American Statistical Association*, 56, 703-728.

——— (1963), "An Investigation of Strategies in Baseball," *Operations Research*, 11, 477-501.

——— (1977), "A Scientific Approach to Strategy in Baseball," in *Optimal Strategies in Sports*, eds. R. E. Machol and S. P. Ladany, Amsterdam: North-Holland, pp. 1-30.

Westfall, P. H. (1990), "Graphical Representation of a Basketball Game," *The American Statistician*, 44, 305-307.