DE GRUYTER

J. Quant. Anal. Sports 2015; 11(3): 145–153

Nicholas G. Polson and Hal S. Stern*

The implied volatility of a sports game

Abstract: In this paper we provide a method for calculating the implied volatility of the outcome of a sports game. We base our analysis on Stern's stochastic model for the evolution of sports scores (Stern, H. S. 1994. "A Brownian Motion Model for the Progress of Sports Scores." *Journal of the American Statistical Association* 89:1128–1134.). Using bettors' point spread and moneyline odds, we extend the model to calculate the market-implied volatility of the game's score. The model can also be used to calculate the time-varying implied volatility during the game using inputs from real-time, online betting and to identify betting opportunities. We illustrate our methodology on data from Super Bowl XLVII between the Baltimore Ravens and the San Francisco 49ers and show how the market-implied volatility of the outcome varied as the game progressed.

Keywords: implied volatility; moneyline odds; point spread; super bowl.

1 Introduction

One of the most exciting aspects of a sporting event is the uncertainty of the outcome. Even when one team is heavily favored, there is a chance that the other team will win and that likelihood varies by the particular sport as well as the particular teams involved. This paper provides an assessment of the uncertainty of the game outcome using as inputs two "market" measures: the betting point spread and the probability of one team winning implied by the betting odds. We calculate the uncertainty measure, the market-implied volatility of the outcome, using the Stern (1994) model for the evolution of the game score. We also show how to update the implied volatility during the game using real-time changes in bettors' assessments that a particular team will win.

We measure the uncertainty of the outcome as the variation associated with the final score of the game. We model the outcome between the two teams as a random process, $X(t)$, which denotes the lead of team A over team B at time t. Negative values indicate that team A is behind. To simplify things, we assume that the game begins at time zero with $X(0) = 0$ and ends at time one with $X(1)$ representing the final score difference (positive if A wins and negative if B wins). The possibility of overtime is not considered. For our analysis we develop a probabilistic specification of the distribution of $X(1)$ and, more generally, of $X(t)$, as the game evolves. Given this probabilistic model we then define the notion of the *implied volatility* of the outcome for the whole game or for the remaining part of the game such as the second half.

Our market-implied game outcome volatility is based on bettors' or analysts' assessments of the game outcomes. In recent years there has been an explosion of online betting markets that provide market-based assessments for the outcomes of many sporting events. The point spread provides an assessment of the *expected margin* of victory, which we will denote by μ. The money-line odds provide an assessment of the *probability that team A wins*, which we denote as p. We show that these are sufficient to define the *implied volatility* which we denote as σ_{IV}. This implied volatility is a market-based assessment of the amount of uncertainty in the difference of scores between the two teams.

The rest of the paper proceeds as follows. Section 2 presents Stern's 1994 model for the evolution of $X(t)$. We then show how the point spread and money line determine the implied volatility of the outcome. We extend this result to calculate a dynamic, time-varying implied volatility using real-time online trading market data for the win probability. Section 2 also discusses limitations and extensions of the basic model. In Section 3 we illustrate our methodology using Super Bowl XLVII between the Baltimore Ravens and the San Francisco 49ers. Finally, Section 4 concludes with directions for future research.

2 The implied volatility of the outcome of a sports game

2.1 A model for sports scores

In order to define the implied volatility of a sports game we begin with a distributional model for the evolution of the

---
*Corresponding author: Hal S. Stern, Department of Statistics, University of California, Irvine, CA, USA, e-mail: sternh@uci.edu
Nicholas G. Polson: University of Chicago Booth School of Business, Chicago, IL, USA

outcome in a sports game which we develop from Stern (1994). The model specifies the distribution of the lead of team A over team B at time t, $X(t)$, as a Brownian motion process. If $B(t)$ denotes a standard Brownian motion with distributional property $B(t) \sim N(0, t)$ and we incorporate terms to accommodate drift, μ, and volatility, σ, then the evolution of the outcome $X(t)$ is given by:

$$X(t) = \mu t + \sigma B(t) \sim N(\mu t, \sigma^2 t).$$

Thus the distribution of the game outcome is a Brownian motion model where $dX_t = \mu dt + \sigma dB_t$. There is a resemblance to the Black-Scholes model of the distribution of a stock price (Hull, 2005) however in that case the underlying model is a geometric Brownian motion. We can use our model, in analogy with the Black-Scholes model for finance, as a means to measure the implied volatility of the outcome; this approach is described in Sections 2.2 and 2.3. Section 2.4 considers limitations of the normal assumption and describes extensions that can include discrete jumps which can be more realistic in certain sports.

The Brownian motion model results in closed-form solutions for a number of probabilities of interest. The distribution of the final score follows a normal distribution, $X(1) \sim N(\mu, \sigma^2)$, where the drift parameter μ measures the advantage in points of team A over the length of the game and the volatility parameter σ measures the standard deviation or volatility of the outcome. We can calculate the probability of team A winning, denoted $p = \mathbb{P}(X(1) > 0)$, from a given point spread (or drift) μ, a given standard deviation (or volatility) σ, and the assumed probability distribution. Given the normality assumption, $X(1) \sim N(\mu, \sigma^2)$, we have

$$p = \mathbb{P}(X(1) > 0) = \Phi\left(\frac{\mu}{\sigma}\right)$$

where Φ is the standard normal cdf. Table 1 uses Φ to convert team A's advantage μ to a probability scale using a few values of the information ratio μ/σ, sometimes also known as the signal-to-noise ratio. The win probability can be easily calculated for any value of the ratio but memorizing the win probability for a few values can be sufficient for identifying betting strategies using the approach discussed below.

If teams are evenly matched and $\mu/\sigma = 0$ then $p = 0.5$. If the point spread $\mu = 4$ (team A is favored by 4 points) and

volatility is $\sigma = 10.6$, then team A has a $\mu/\sigma = 4/10.6 = 0.38$ volatility point advantage. The probability of winning is $\Phi(0.38) = 0.65$. An informative discussion of how the point spread is set in practice is provided in Stern (1999).

Of particular interest here are conditional probability assessments made as the game progresses. For example, suppose that the current lead at time t is l points and so $X(t) = l$. The model can then be used to update your assessment of the distribution of the final score with the conditional distribution $(X(1) | X(t) = l)$. We can derive the conditional distribution of $X(1)$ given $X(t)$ by noting that $X(1) = X(t) + (X(1) - X(t))$. As $X(t) = \mu t + \sigma B(t)$ and under the conditioning $X(t) = l$, this simplifies to

$$X(1) = l + \mu(1-t) + \sigma(B(1) - B(t)).$$

Here $B(1) - B(t) \overset{D}{=} B(1-t)$ which is independent of $X(t)$ with distribution $N(0, 1-t)$.

To determine the conditional distribution, we first note that there are $1 - t$ time units left in the game together with a drift μ so that the conditional mean is the current lead l plus the remaining expected advantage (or disadvantage) associated with the drift $\mu(1-t)$. The conditional uncertainty can be modeled as $\sigma B(1-t)$ which is $N(0, \sigma^2(1-t))$. Therefore, we can write the distribution of the final outcome, $X(1)$, given that time t has elapsed and team A leads by l points as the conditional distribution:

$$X(1) | X(t) = l \sim N(l + \mu(1-t), \sigma^2(1-t))$$

From the conditional distribution we can calculate the conditional probability of winning as the game evolves. The probability of team A winning at time t given a current lead of l points is:

$$p_{t,l} = P(X(1) > 0 | X(t) = l) = \Phi\left(\frac{l + \mu(1-t)}{\sigma\sqrt{(1-t)}}\right)$$

where we have now subscripted p to denote the conditioning on the time and lead.

The conditional distribution of the outcome given the score at time t has variance $\sigma^2(1-t)$ or volatility of $\sigma\sqrt{1-t}$. The volatility is a decreasing function of t, illustrating that the volatility dissipates over the course of a game. For example, if there is an initial volatility of $\sigma = 10.6$ points, then at half-time when $t = 0.5$, the remaining volatility is $10.6/\sqrt{2} = 7.5$ points. Table 2, below, illustrates this relationship for several time points over the course of the game assuming the initial volatility is 10.6 points.

Tables 1 and 2 can be combined to measure the significance of the current outcome, l, in terms of the remaining volatility in the game. We demonstrate with an example (motivated by Section 3) for which it is more natural to

Table 1: Probability of winning p versus the ratio μ/σ.

μ/σ	0	0.25	0.5	0.75	1	1.25	1.5	2
$p = \Phi(\mu/\sigma)$	0.50	0.60	0.69	0.77	0.84	0.89	0.93	0.98

Table 2: Volatility decay over time.

t	0	0.25	0.50	0.75	1
$\sigma\sqrt{1-t}$	10.6	9.2	7.5	5.3	0

consider the perspective of the underdog team B. Suppose that you have Team B, a four-point underdog, so from their perspective $\mu = -4$ and at half-time team B has a lead of 15, $l = 15$. Team B's expected outcome starting from these conditions is $l + \mu(1-t)$ or $15 - 4 \times 0.5 = 13$. If the initial volatility is $\sigma = 10.60$, then the remaining volatility at half-time is $10.6/\sqrt{2} = 7.50$, and team B's expected outcome of 13 is $13/7.5 = 1.73$ standard deviations above zero. Thus team B's expected outcome is at the 96th percentile of its distribution, $\Phi(1.73) = 0.96$, implying a 96% chance of winning the game.

Figure 1 illustrates how the different aspects of the model that we have discussed can be visualized with an example. Suppose we are analyzing data for the football game between teams A and B described above with team A favored by 4 points initially and with volatility 10.6 points. The top left panel of the figure demonstrates the

information available at the beginning of the game from the perspective of the underdog team B. The outcome distribution from the perspective of team B under our normal assumption is $X(1) \sim N(-4, 10.6^2)$. This distribution is illustrated by the density curve in the figure and the probability that B wins is the area of the shaded red area under the curve (0.35). The top right panel illustrates a simulation-based approach to visualizing how the model works. Here we show a set of simulated Brownian motion paths with the given volatility. Roughly 35% of the simulations end up with Team B ahead. The bottom row shows the same two figures using information available at half-time. The bottom left panel shows the outcome distribution given the conditional mean $(l + \mu(1-t) = 15 - 4 \times 0.5 = 13)$ and standard deviation $(10.6/\sqrt{2} = 7.5)$. The updated win probability is now the area of the large shaded red area. The bottom right panel shows the evolution of the actual score until half-time as the solid black line. From half-time onwards we simulate a set of possible Monte Carlo paths to the end of the game. All but a few of the paths now show team B winning. The volatility plays a key role in turning the point spread into a probability of winning as the greater the volatility of the distribution of the outcome, $X(1)$, the greater the range of outcomes projected in the
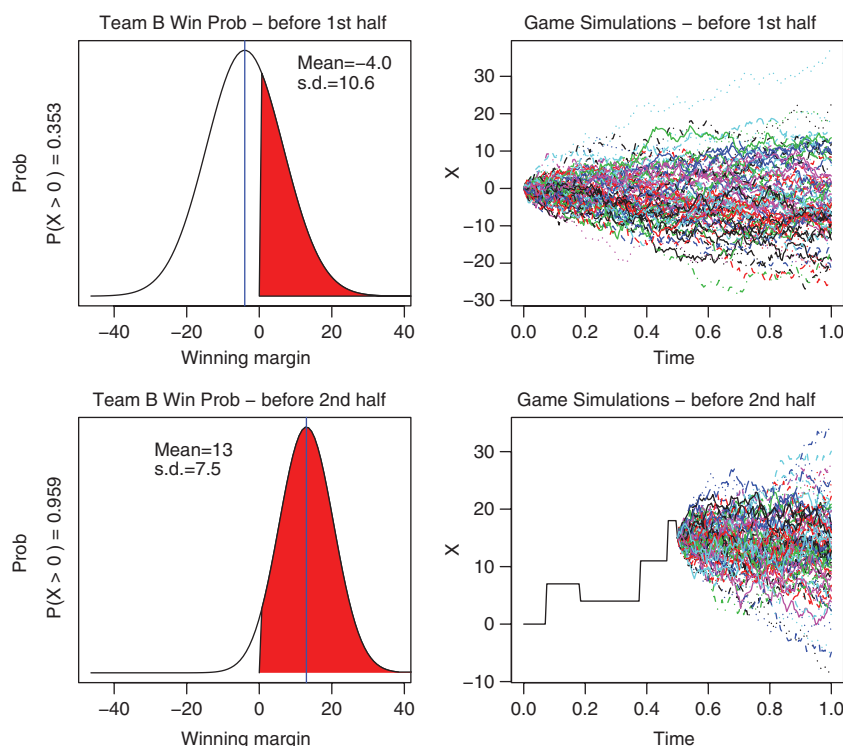
Figure 1: Illustration of the Brownian motion model. The top left panel shows the outcome distribution using information before the game starts and the top right panel shows a set of simulated Brownian motion paths for the game outcome. The bottom row has the two figures updated using information available at half-time.

Monte Carlo simulation. Essentially the volatility provides a scale which calibrates the advantage implied by a given lead at a given time.

2.2 Implied volatility

The previous discussion assumed that the variance (or volatility) parameter σ was a known constant. We return to this important quantity now. We are now in a position to define the *implied volatility* implicit in the two betting markets (point spread and money line) that are available. Given our model, we will use the *money-line* odds to provide a market assessment of the probability of winning, p, and the *point spread* to assess the expected margin of victory, μ. The money-line odds are given for each team and provide information on the payoff from a bet on that team winning. The payoff can be used to infer the market's assessment of the probability that the team will win. As shown in the example in Section 3, this calculation will also typically require an adjustment for the bookmaker's spread. Once we have μ and p, we can infer the *implied volatility*, σ_{IV}, by solving

$$p = \Phi\left(\frac{\mu}{\sigma_{IV}}\right) \text{ which gives } \sigma_{IV} = \frac{\mu}{\Phi^{-1}(p)} \ .$$

Here $\Phi^{-1}(p)$ denotes the standard normal quantile function such that the area under the standard normal curve to the left of $\Phi^{-1}(p)$ is equal to p. In our example we calculate this using the `qnorm` function in R. Note that when $\mu = 0$ and $p = 0.5$ there's no market information about the volatility as $\mu/\Phi^{-1}(p)$ is undefined. This is the special case where the teams are seen as evenly matched – the expected outcome has a zero point spread and there is an equal probability that either team wins.

2.3 Time-varying implied volatility

Up to this point the volatility rate has been assumed constant through the course of the game, i.e., that the same value of σ is relevant at all times. The amount of volatility remaining in the game is not constant because time has elapsed but the basic underlying parameter has been assumed constant. This need not be true and more importantly the betting markets may provide some information about the best estimate of the volatility parameter at a given point in time. This is important because time-varying volatility provides an interpretable quantity that can allow one to assess the value of a betting opportunity.

With the advent of on-line betting there is a virtually continuously traded contract available to assess the implied expectation of the probability of team A winning the game at any time t. The additional information available from the continuous contract allows for an update of the implied volatility conditional on the game situation at time t. We assume that the online betting market gives us a current assessment of $p_{t,l}$, the probability that team A will win based on all information available at time t. We can use our previously defined expression for the conditional probability of A winning to solve for an estimate of the time-varying implied volatility, $\sigma_{IV,t}$:

$$p_{t,l} = \Phi\left(\frac{l + \mu(1-t)}{\sigma_{IV,t}\sqrt{1-t}}\right) \text{ which gives } \sigma_{IV,t} = \frac{l + \mu(1-t)}{\Phi^{-1}(p_{t,l})\sqrt{1-t}}$$

where the notation indicates that the conditional win probability $p_{t,l}$ depends on the current time and lead.

This analysis assumes that the drift parameter μ remains constant and does not vary over time. Given only a single probability assessment at time t it is not possible to infer time-varying values for both parameters. There are some instances in which it may seem more natural to allow μ to vary but by and large we believe it is more natural to expect that the underlying volatility, σ, may vary because of game conditions such as the weather and current score.

We will use our methodology in the next Section to find evidence of time-varying volatility during Super Bowl XLVII.

2.4 Model extensions

One of the main assumptions in the model we are using is the normality assumption that underlies the Brownian motion. The model was initially proposed by Stern (1994) for basketball, a high-scoring sport for which the normal distribution is a plausible starting point. Even though Rosenfeld (2012) finds evidence of heavier than normal tails throughout the game and non-normality of score changes at the end of the game. Rosenfeld (2012) provides an extension of the model to address these concerns; the normal distribution is replaced by a logistic distribution and the relative contribution of the lead and the remaining advantage are estimated empirically. The normality assumption is even more questionable in low-scoring sports such as baseball, hockey or soccer which are categorized by a series of discrete scoring events. Football, the sport of interest here, is somewhere in between these two extremes. The scores are large enough to support a

normal approximation for the margin of victory (see, for example, Stern (1991)) but discrete scoring events (e.g., touchdowns) have a large impact on the score, especially near the end of the game.

One approach to addressing the presence of discrete scoring events is to model the interarrival times. Thomas (2007) carries out a study of this type for hockey and Chu (2003) does the same for soccer. These models can be used to develop a Poisson process model of the game outcome. We now show how to add a Poisson jump component to a Brownian motion model in order to allow for discrete shocks to the expected score difference between the teams. The equivalent in finance is Merton's (1976) jump model. Specifically, we model

$$dX_t = \mu dt + \sigma dB_t + \xi_t dJ_t$$

where $\xi_t \sim g(\theta_J)$, $g(\theta_J)$ is a distribution depending on parameters θ_J that governs the size of the jumps, and J_t is a Jump process with the rate of jumps governed by the parameter λ (i.e., $\mathbb{P}(dJ_t = 1) = \lambda dt$). The nature of the jump distribution depends on the sport. For example, in football it could be a distribution that accounts for the possibility of scores of size ± 2, ± 3, and ± 7. We now have a model parameterised by $(\mu, \sigma, \lambda, \theta_J)$ and with only two pieces of market information (the point spread and money line) we have an ill-posed inverse problem. This makes it harder to implement than the Brownian motion model; we either have to subjectively assess the jump component parameters (θ_J, λ) or estimate them from historical data before calibrating (μ, σ) to the betting line. One interesting feature of this model is that the jump component is added to the *difference in score* of the two teams.

An alternative approach is to develop an appropriate probability model directly on the game scores. For example, one might propose a bivariate Poisson distribution for the number of goals/points scored for each team in hockey or soccer; see, for example, theoretical results of Keller (1994), Dixon and Coles (1997) and empirical analyses of Lee (1997), Kaslis and Ntzoufras (2003), Speigelhalter and Ng (2009). Our approach applies equally well there and we can infer the implied mean number of goals to be scored given the probability associated with a betting contract.

We have only considered bettors' odds as providing information about the drift parameter μ and the volatility parameter σ. As in the financial markets an alternative is to use historical data to estimate one or both of these parameters. Stern (1991) provides empirical evidence regarding the value of σ. There are numerous models that have been proposed for using historical data to estimate the advantage of one team over another (μ), see, for

example, Glickman and Stern (1998) and the references therein.

3 Super Bowl XLVII: Baltimore Ravens vs. San Francisco 49ers

Super Bowl XLVII was held at the Superdome in New Orleans on February 3, 2013 and featured the San Francisco 49ers against the Baltimore Ravens. Going into Super Bowl XLVII the San Francisco 49ers were favorites to win which was not surprising following their impressive season. The Ravens opened up a substantial lead, 28–6, early in the third quarter. This was followed by a rather unusual event – the stadium suffered a 34 min power outage. When power was restored the 49ers began to come back. The game's outcome was in doubt late in the final quarter when the 49ers were close to scoring a go-ahead touchdown. The 49ers failed to score however and the Ravens held on for a 34–31 upset victory.

We applied the model of Section 2 from the viewpoint of the Ravens so that $X(t)$ corresponds to the Raven's lead over the 49ers at time t. Table 3 provides the score at the end of each quarter.

The parameters of our model were obtained using information from the pregame betting markets. The initial *point spread* established the Ravens as a 4 point underdog so we set the mean of our outcome, $X(1)$, as

$$\mu = \mathbb{E}(X(1)) = -4.$$

The markets' pregame assessment of the probability that the Ravens would win is determined from the pregame *money-line* odds. These odds were quoted as San Francisco −175 and Baltimore +155. This means that a bettor would have to bet $175 to win $100 on the 49ers while a bet of $100 on the Ravens would lead to a win of $155. We converted both of these money lines to *implied probabilities* of each team winning, using the equations

$$p_{SF} = \frac{175}{100 + 175} = 0.686 \text{ and } p_{Bal} = \frac{100}{100 + 155} = 0.392$$

Table 3: Super Bowl XLVII by quarter.

t	0	0.25	0.50	0.75	1
Ravens	0	7	21	28	34
49ers	0	3	6	23	31
$X(t)$	0	4	15	5	3

The resulting probabilities do not sum to one. This "excess" probability is in fact the mechanism by which the oddsmakers derive their compensation. The probabilities actually sum to one plus a quantity known as the "market vig", also known as the bookmaker's edge. In this example,

$$p_{SF} + p_{Bal} = 0.686 + 0.392 = 1.078$$

providing a 7.8% edge for the bookmakers. Put differently, if bettors place money proportionally across the two teams then the bookies will make 7.8% of the total staked no matter the outcome of the game. To account for this edge in our model, we used the mid-point of the two implied probabilities that Baltimore would win to determine p. This yielded

$$p = \frac{1}{2} p_{Bal} + \frac{1}{2}(1 - p_{SF}) = 0.353$$

Thus, from the Ravens perspective we took the initial probability of a win to be $p = \mathbb{P}(X(1) > 0) = 0.353$.

Figure 2 shows the evolution of the markets' assessment of the conditional probability of the Ravens winning as the game progresses, which we denote by p_t^{mkt}. These probabilities are derived from betting prices obtained from the online betting website `TradeSports.com`. The contract volumes (amount of betting) are displayed along the x-axis. Baltimore's win probability started

trading at $p_0^{mkt} = 0.38$ (similar to the initial estimate derived from the money odds above) and the dynamic probabilities fluctuated dramatically throughout the course of an exciting game. The Ravens took a commanding 21–6 lead at half time and the market was trading at $p_{0.5}^{mkt} \approx 0.90$. The market-implied probability increased to 0.95 when Baltimore extended its lead by returning the second-half kickoff for a touchdown. This was followed by the power blackout. During the blackout 42,760 contracts changed hands with Baltimore's win probability ticking down from 0.95 to 0.94. There was considerable movement in the market probability as the game became closer in the fourth quarter. Near the end of the fourth quarter, the 49ers had the ball in position to take the lead (the point labelled "49ers 1st & Goal" in the figure) and Baltimore's win probability had dropped to just above 30%. The 49ers failed to score in the course of the next four plays and the win probability jumped back up near one.

Our model allows us to provide answers for a number of important questions:

What implied volatility is consistent with pregame market expectations?
The implied volatility of the Super Bowl is calculated by substituting the pair $(\mu, p) = (-4, 0.353)$ into our definition and solving for σ_{IV}. We obtained
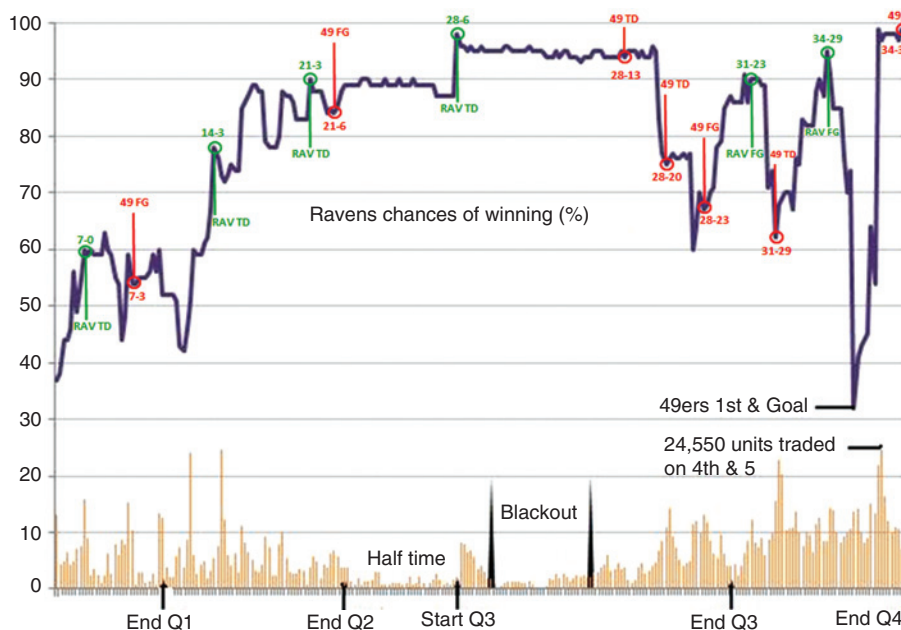
Figure 2: Betting data for Super Bowl XLVII between the Ravens and 49ers obtained from the `TradeSports.com` website. The solid line shows the dynamic market probabilities of the Ravens winning with the vertical axis giving the probability scale (numbers expressed as percentages). The number of contracts traded at each point in time is represented by the vertical bars along the bottom of the figure; these can also be interpreted with reference to the vertical axis with the numbers now representing thousands of units traded.

$$\sigma_{IV} = \frac{\mu}{\Phi^{-1}(p)} = \frac{-4}{-0.377} = 10.60$$

(the value we have been using throughout the article) where we have used $\Phi^{-1}(p) = qnorm(0.353) = -0.377$. So on a volatility scale the 4 point pregame advantage assessed for the 49ers is less than $0.5\,\sigma$. The inferred $\sigma = 10.6$ is historically low, as the typical volatility of an NFL game is somewhere between 13 and 14 points (see Stern, 1991). One possible explanation is that for a competitive game like this, one might expect a lower than usual volatility. (Of course another possibility is that the two market assessments, the point spread and the win probability, are inconsistent for some reason.) The outcome $X(1) = 3$ was within one standard deviation of the pregame model which had an expectation of $\mu = -4$ and volatility of $\sigma = 10.6$. As the game progresses our framework can be used to address a variety of additional questions.

What's the probability of the Ravens winning given their lead at half time?
At half time Baltimore led by 15 points, 21 to 6. There are two ways to estimate the probability that Baltimore will win the game given this half time lead. Stern (1994) assumes the volatility parameter is constant during the course of the game (and equal to the pregame value). Using this assumption we applied the formula derived earlier with conditional mean for the final outcome $15 + 0.5*(-4) = 13$ and conditional volatility $10.6\sqrt{1-0.5} = 7.5$. These yielded a probability of .96 for Baltimore to win the game.

A second estimate of the probability of Baltimore winning given their half time lead can be obtained directly from the betting market. From the online betting market we have traded contracts on `TradeSports.com` that yield a half time probability of $p_{0.5}^{mkt} = 0.90$. There is a notable difference in the two estimates. One possible explanation for the difference is that the market assesses time varying volatility and the market price (probability) reflects a more accurate underlying probability. This leads to further study of the implied volatility.

What's the implied volatility for the second half of the game?
We determined the market's assessment of implied volatility at half time by assuming that p_t^{mkt} reflects all available information. We applied the formula of Section 2.3 using $t = 0.5$ and obtained

$$\sigma_{IV, t=0.5} = \frac{l + \mu(1-t)}{\Phi^{-1}(p_t)\sqrt{1-t}} = \frac{15-2}{\Phi^{-1}(0.9)\sqrt{0.5}} = 14$$

where $qnorm(0.9) = 1.28$. As $14 > 10.6$, the market was expecting a more volatile second half – possibly anticipating a comeback from the 49ers. It is interesting to note that the implied half time volatility parameter is more consistent with the value typical for regular season games.

How can we use this framework to form a valid betting strategy?
The model provides a method for identifying good betting opportunities. If you believe that the initial money line and point spread identify an appropriate value for the volatility (10.6 points in the Super Bowl example) – and this stays constant throughout the game – opportunities arise when there's a difference between the dynamic market probabilities and the probabilites obtained from the Brownian motion model. This can be a reasonable assumption unless there has been some new material information such as a key injury. For example, given the initial implied volatility $\sigma = 10.6$, at half time with the Ravens having a $l + \mu(1-t) = 13$ points edge we would assess a

$$p_{0.5} = \Phi\left(13 / (10.6 / \sqrt{2})\right) = 0.96$$

probability of winning versus the $p_{0.5}^{mkt} = 0.90$ rate. This suggests a bet on the Ravens is in order. To determine our optimal bet size, ω_{bet}, we might appeal to the Kelly criterion (Kelly, 1956) which would yield

$$\omega_{bet} = p_{0.5} - \frac{q_{0.5}}{O^{mkt}} = 0.96 - \frac{0.1}{1/9} = 0.60$$

where $p_{0.5}$ is our assessment of the probability of winning at half time (0.96), $q_{0.5} = 1 - p_{0.5}$ and the market offered odds are given by

$$O^{mkt} = (1 - p^{mkt}) / p^{mkt} = (1-0.9) / 0.9 = 1 / 9.$$

This would imply a bet of 60% of capital. The Kelly criterion is the optimal bet size for long-term capital appreciation. In this setting the bet size seems quite high. It is mathematically correct since we believe we have a 96% chance of earning a 10% return but the bet size is very sensitive to our estimated probability of winning. Given that the Brownian motion model is just an approximation we may wish to be more conservative. It is common in such situations to bet less than the optimal Kelly bet size. The fractional Kelly criterion scales one's bet by a risk aversion parameter, γ, often $\gamma = 2$ or $\gamma = 3$. Here with $\gamma = 3$, the bet size would then be $0.60/3 = 0.20$, or 20% of capital.

Alternate strategies for assessing betting opportunities are possible. The assumption of constant volatility in the previous paragraph is often reasonable but not always. There may be new information available such as a change in weather conditions or an expected change in team strategy (e.g., given the deficit the 49ers might be expected to take more chances). If so, the implied volatility calculation demonstrated above provides a useful way for bettors to evaluate their opportunities. In the Super Bowl one might judge that the half-time implied volatility (14 points) is in fact more realistic than the pre-game value (10.6 points). This would lead one to conclude that the market probability $p_{0.5}^{mkt}$ is the more accurate estimate and there would not be a betting opportunity at half time of the game.

4 Discussion

By combining information from Stern's Brownian motion model for the evolution of sports scores and information from betting and prediction markets we defined the implied volatility of a sporting event. The explosion of online betting and prediction markets provides an opportunity to learn from market-based information through the lens of our probabilistic model. There are two questions of interest in assessing our approach. First, is the model reasonable? and, second, how informative is the market information?

There are a number of features and assumptions of the model that can be relaxed. Section 2.4 discussed the normal distribution assumption. Another key assumption is the Markov nature of the probabilistic information that is used. Is it really the case that the only relevant conditioning information is the current lead ($X(t)=l$) and that the earlier history in the game is irrelevant? Interestingly, this was one of the original criticisms of the Black-Scholes model. However, in a competitive sporting situation one could imagine that this is a more realistic assumption than many people might initially think. Of course, there are exceptions. For example, Jackson (1993) considers the non-Markovian nature of tennis games and the effect of dependence on the assessment of rare event probabilities.

There have been many studies of the informational content of betting and prediction markets. See Snowberg, Wolfers, and Zitewitz (2012) for a recent discussion. We briefly review some relevant literature here. Avery and Chevalier (1999) ask whether you can identify sentiment from price paths in football betting. In many markets this is the issue of noise traders who are not necessarily providing any extra new information to the bettor. Camerer (1989) asks whether the point spread in basketball provides evidence of whether the market believes in the hot hand. He finds that extreme underdogs, teams that have been on a long losing streak, are under-priced by the market. There are also other forms of market information that are useful for measuring volatility. One area to analyze are index betting spreads (Jackson, 1994) where bettors provide assessments of over-under lines. These lines are used to make bets about the total number of points scored in a game. For example, in Super Bowl XLVII the over-under line was 48 with the true total number of points being $34 + 31 = 65 > 48$. The line 48 provides information about how high the market thinks the score will be. In the future, there might be other betting markets that open up. The win market is essentially a binary option on the outcome with a zero-one payout depending on who wins. Introducing bets with payouts that are proportional to the magnitude of the victory would provide markets more sensitive to the volatility of the game and would be expected to provide more information about volatility.

Our focus on implied volatility provides an interesting perspective on the relationship between the point spread and the money-line odds. There is not a one-to-one mapping between the point spread and money-line odds as is commonly thought. For example, it is not uncommon in college basketball to have two games both with a heavy favorite where the money-line odds are the same but the point spread in one game is significantly higher than the point spread in the other game. Our explanation for this difference is a simple one – the market has an expectation that the volatility is much higher for the game with the larger point spread.

We should remark that as with the Black-Scholes model for option prices in financial markets, our definition of implied volatility does not rely on our model being true *per se*. Implied volatility is merely a lens through which we can communicate bettors' information about current expectations for the future course of a sporting event.

References

Avery, C. and J. Chevalier. 1999. "Identifying Investor Sentiment from Price Paths: The Case of Football Betting." *Journal of Business* 72(4):493–521.

Camerer, C. F. 1989. "Does the Basketball Market Believe in the Hot Hand?" *American Economic Review* 79(5):1257–1261.

Chu, S. 2003. "Using Soccer Goals to Motivate the Poisson Process." *INFORMS Transactions on Education* 3(2):64–70.

Dixon, M. J. and S. G. Coles. 1997. "Modelling Association Football Scores and Inefficiencies in the Football Betting Market." *Applied Statistics* 46:265–280.

Glickman, M. E. and H. S. Stern. 1998. "A State-Space Model for National Football League Scores." *Journal of the Amercan Statistical Association* 93:25–35.

Hull, J. C. 2005. *Options, Futures and Other Derivatives*. 6th ed. New Jersey: Prentice Hall.

Jackson, D. A. 1993. "Independent Trails are a Model for Disaster." *Journal of the Royal Statistical Society. Series C* 42:211–220.

Jackson, D. A. 1994. "Index Betting on Sports." *Journal of the Royal Statistical Society. Series D* 43(2):309–315.

Kaslis, D. and I. Ntzoufras. 2003. "Analysis of Sports Data by using Bivariate Poisson Models." *The Statistician* 52(3):381–393.

Keller, J. B. 1994. "A Characterisation of the Poisson Distribution and the Probability of Winning a Game." *The American Statistician* 48(4):294–298.

Kelly, J. L. 1956. "A New Interpretation of the Information Rate." *Bell System Technical Journal* 35(4):917–926.

Lee, A. J. 1997. "Modeling Scores in the Premier League: Is Manchester United Really the Best?" *Chance* 10(1): 15–19.

Merton, R. C. 1976. "Option Pricing when Underlying Stock Returns are Discontinuous." *Journal of Financial Economics* 3:125–144.

Rosenfeld, J. W. 2012. "An In-game Win Probability Model of the NBA." *Thesis, Harvard University*.

Snowberg, E., J. Wolfers, and E. Zitewitz. 2012. "Prediction Markets for Economic Forecasting." *Working Paper*.

Speigelhalter, D. S. and Y-L. Ng. 2009. "One Match to Go!" *Significance* 6(4):151–154.

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

Stern, H. S. 1994. "A Brownian Motion Model for the Progress of Sports Scores." *Journal of the American Statistical Association* 89:1128–1134.

Stern, H. S. 1999. "The Man who Makes the Odds: An Interview with Roxy Roxborough." *Chance* 12(1):15–21.

Thomas, A. C. 2007. "Inter-arrival times of goals in Ice Hockey." *Journal of Quantitative Analysis in Sports* 3(3):5.