

The Implied Volatility of a Sports Game

Nicholas G. Polson

Hal S. Stern

Booth School of Business

UC Irvine*

First Draft: August 2013

This Draft: March 2014

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, 1994). 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, Point Spread, Moneyline Odds, Super Bowl.

**Address:* Nick Polson is at the University of Chicago Booth School of Business. email: ngp@chicagobooth.edu. Hal Stern is at UC Irvine. email: sternh@uci.edu. We thank two referees and the Associate Editor for their comments.

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 as measured by a “market” assessment using the point spread and probability of one team winning typically quoted by betting markets or sports pundits as inputs. We calculate this measure using Stern’s 1994 model for the evolution of the game score. We extend this result to estimate the market-implied volatility of the outcome and show how to update the volatility during the game with 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 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 $X(1)$ determining the final score. 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 the remainder 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 moneyline 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 game time implied volatility using real-time online trading market data for the win probability. In section 3 we illustrate our methodology using SuperBowl XLVII between the Baltimore Ravens and the San Francisco 49ers. Finally, Section 4 concludes with directions for future research.

2 The Implied Volatility of a 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 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, $X(t)$ for any 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 drift, μ , and volatility, σ , terms, then the evolution of the outcome $X(t)$ is given by:

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

This distribution of the game outcome is a Brownian motion model where $dX_t = \mu dt + \sigma dB_t$ rather than the geometric Brownian motion model used in the Black-Scholes model of the distribution of a stock price (Hull, 2005). However, in a similar manner we can use our model as a lens to measure the implied volatility of the outcome as we describe in Section 2.2. Section 2.4 considers extensions that include discrete jumps which can be more realistic in certain sports.

This specification results in closed-form solutions for a number of probabilities of interest. The distribution of the final score follows a normal distribution, $X(1) \sim N(\mu, \sigma^2)$. 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 the information ratio μ/σ .

μ/σ	0	0.25	0.5	0.75	1	1.25	1.5	2
$p = \Phi(\mu/\sigma)$	0.5	0.60	0.69	0.77	0.84	0.89	0.93	0.977

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

If teams are evenly matched and $\mu/\sigma = 0$ then $p = 0.5$. Table 1 provides a list of probabilities as a function of μ/σ . For example, if the point spread $\mu = -4$ and volatility is $\sigma = 10.6$, then team A has a $\mu/\sigma = -4/10.6 = -0.38$ volatility point disadvantage. The probability of winning is $\Phi(-0.38) = 0.353 < 0.5$. A common scenario is that team A has an edge equal to half a volatility, in that case $\mu/\sigma = 0.5$ and then $p = 0.69$.

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 $(X(1)|X(t) = l)$. To see this, we can re-write the 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 $B(1) - B(t) \stackrel{D}{=} B(1 - t)$ 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 after time t has elapsed with a current lead of l for team A as the conditional distribution:

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

From the conditional distribution $(X(1)|X(t) = l) \sim N(l + \mu(1-t), \sigma^2(1-t))$, 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:

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

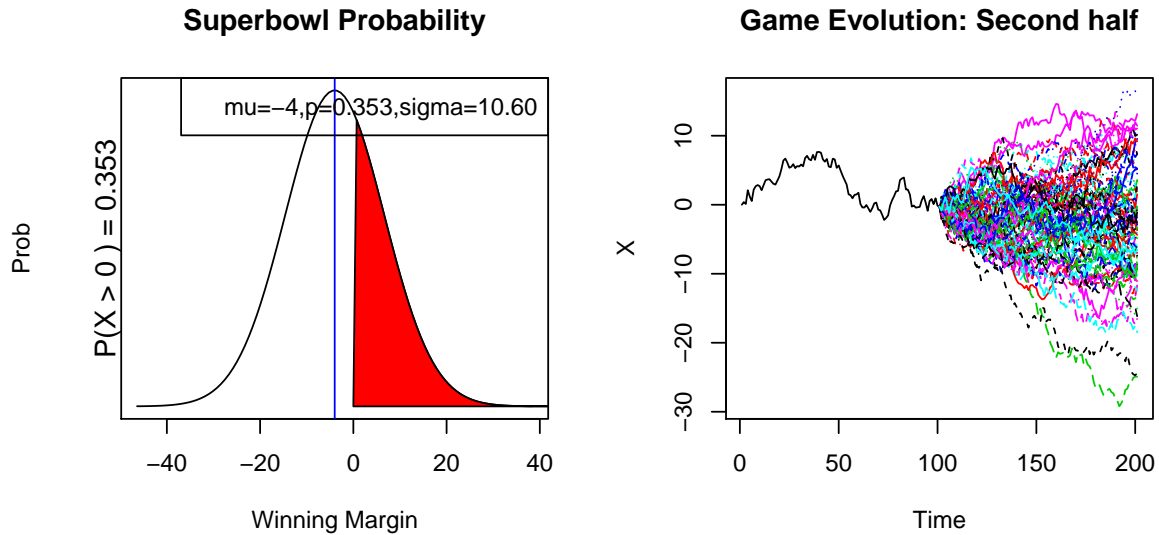


Figure 1: Score Evolution on a Discretised Grid

Figure 1 illustrates our methodology with an example. Suppose we are analyzing data for a Super Bowl football game between teams A and B with team A favored. The left panel presents the information available at the beginning of the game from the perspective of the underdog team B. If the initial point spread—or the markets’ expectation of the expected outcome—is -4 and

the volatility is 10.6—assumed given for the moment (more on this below)—then the probability that the underdog team wins is $p = \Phi(\mu/\sigma) = \Phi(-4/10.6) = 35.3\%$. This result relies on our assumption of a normal outcome distribution on the outcome as previously explained. Another way of saying this is $\mathbb{P}(X(1) > 0) = 0.353$ for an outcome distribution $X(1) \sim N(-4, 10.6^2)$. This is illustrated with the shaded red area under the curve.

The right panel illustrates the information and potential outcomes at half-time. Here we show 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. 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 Monte Carlo simulation. Essentially the volatility provides a scale which calibrates the advantage implied by a given point spread.

We can use this relationship to determine how volatility decays over the course of the game. The conditional distribution of the outcome given the score at time t , is $(X(1)|X(t) = l)$ with a variance of $\sigma^2(1 - t)$ and 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 $\sigma = 10.6$, then at half-time when $t = \frac{1}{2}$, the volatility is $10.6/\sqrt{2} = 7.5$ volatility points left. Table 2, below, illustrates this relationship for additional points over the game.

t	0	$\frac{1}{4}$	$\frac{1}{2}$	$\frac{3}{4}$	1
$\sigma\sqrt{1 - t}$	10.6	9.18	7.50	5.3	0

Table 2: Volatility Decay over Time

To provide insight into the final outcome given the current score, Tables 1 and 2 can be combined to measure the current outcome, l , in terms of standard deviations of the outcome. For example, suppose that you have Team B, an 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 as presented earlier is $l + \mu(1 - t)$ or $15 - 4 \times \frac{1}{2} = 13$. If 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 in terms of standard deviations is $13/7.5 = 1.73$. Thus team B's expected outcome is at the 96th percentile of the distribution, as $\Phi(1.73) = 0.96$, implying a 96% chance of winning.

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 lines 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 shown for each team A and B and provide information on the payoff from a bet on the team winning. As shown in the example in section 3, this calculation will also typically require an adjustment for the bookmaker's spread. With these we can infer the *implied volatility*, σ_{IV} , by solving

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

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` in R. Note that when $\mu = 0$ and $p = \frac{1}{2}$ 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. 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 of 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 implied expectations of the probability of team A winning at any time t . The additional information available from the continuous contract allows for further update of the implied conditional volatility. We assume that the online betting market gives us a current assessment of p_t , that is the current probability that team A will win. 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, $\sigma_{IV,t}$:

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

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

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

2.4 Model Extensions

One of the main assumptions in the Stern model is the Brownian motion of normality assumption. This is not realistic in sports such as baseball or soccer. For example, Thomas (2007) provides an analysis of the inter-arrival times between goals in hockey. We now show how to add an additional jump component 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

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

where $\xi_t \sim \mathcal{N}(\mu_J, \sigma_J^2)$ and J_t is a Jump process where $\mathbb{P}(dJ_t = 1) = \lambda dt$ with a rate of jumps governed by λ . We now have a five parameter model parameterised by $(\mu, \sigma, \lambda, \mu_J, \sigma_J)$ and with only two pieces of market information (the point spread and moneyline) we have an ill-posed inverse problem. This makes it harder to implement than the vanilla Brownian motion model and we either have to subjectively assess the jump component parameters $(\mu_J, \sigma_J, \lambda)$ or estimate them from historical data before calibrating (μ, σ) to the betting line. The win probability can be computed by noting that there's a Poisson number of jumps, $J \sim \text{Poi}(\lambda)$ implied by this model and by the law of iterated expectation

$$\mathbb{P}(X(1) > 0) = \mathbb{E}_J(\mathbb{P}(X(1) > 0|J)) \quad \text{where} \quad \mathbb{P}(X(1) > 0|J) = \Phi\left(\frac{\mu + J\mu_J}{\sqrt{\sigma^2 + J\sigma_J^2}}\right).$$

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 a bivariate Poisson distribution for the number of goals/points scored for each team; see, for example, theoretical results of Keller (1994), Dixon and Coles (1997) and empirical analysis of Lee (1997), Kaslis and Ntzoufras (2003), Spiegelhalter and Ng (2009). Our approach applies equally well here and we can infer the implied mean number of goals to be scored given the probability associated with a betting contract.

We have only consider using a model to provide an implied volatility from the bettors' odds-making. An informative discussion of how the point spread markets are set in practice is provided in Stern (1999). Another area of statistical interest is in models that provide assessments of point spreads from historical data, see, for example, Glickman and Stern (1998). When estimation is the main aim, there is a clear need to allow for more flexible statistical modeling. Rosenfeld (2012) provides a detailed extension of the Stern model for updating in-game win probabilities and tackles the interesting problem of end-of-game assessment where the normality assumption is least palatable.

3 Super Bowl XLVII: 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. It was a fairly bizarre Super Bowl with a 34 minute power outage affecting the game and ultimately an exciting finish with the Ravens causing an upset victory 34 – 31. We will build our model from the viewpoint of the Ravens. We will track $X(t)$ which corresponds to the Raven’s lead over the 49ers at each point in time. Table 3 provides the score at the end of each quarter.

t	0	$\frac{1}{4}$	$\frac{1}{2}$	$\frac{3}{4}$	1
<i>Ravens</i>	0	7	21	28	34
<i>49ers</i>	0	3	6	23	31
$X(t)$	0	4	15	5	3

Table 3: SuperBowl XLVII by Quarter

To determine the parameters of our model we first use the initial *point spread* which was set at the Ravens being a four point underdog, $\mu = -4$. This sets the mean of our outcome, $X(1)$, as

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

In reality, it was an exciting game with the Ravens upsetting the 49ers by 34 – 31. Hence, the realized outcome is $X(1) = 34 - 31 = 3$ with the point spread being beaten by 7 points.

To determine the markets’ assessment of the probability that the Ravens would win at the beginning of the game we use the *money-line* odds. These odds were quoted as San Francisco –175 and Baltimore Ravens +155. This implies that a bettor would have to place \$175 to win \$100 on the 49ers and a bet of \$100 on the Ravens would lead to a win of \$155. We can convert both of these money-lines to *implied probabilities* of the each team winning, by the equations

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

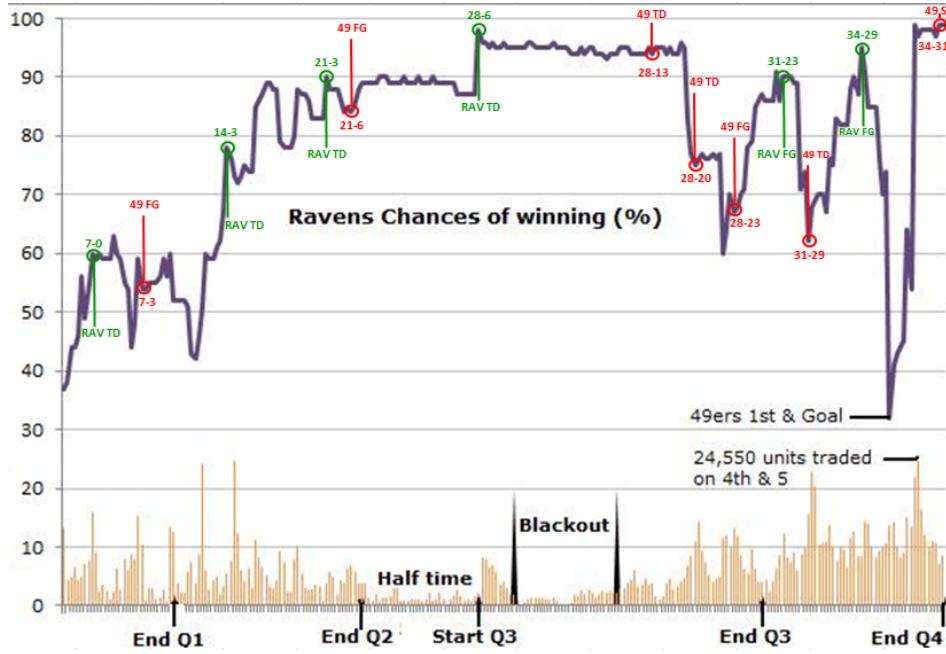


Figure 2: Superbowl XLVII: Ravens vs 49ers: TradeSports contracts traded and dynamic market probabilities, p_t^{mkt} , of the Ravens winning

The 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 both 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 use the mid-point of the spread to determine p implying that

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

From the Ravens perspective we have the initial probability of a win $p = \mathbb{P}(X(1) > 0) = 0.353$.

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 prices are 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 $p_{\frac{1}{2}}^{mkt} = 0.90$. During the 34 minute black-out 42760 contracts changed hands with Baltimore's win probability ticking down from 95 to 94. The win probability peak of 95% occurred again after a third-quarter kickoff return for a touchdown. At the end of the four quarter, however, when the 49ers nearly went into the lead with a touchdown, Baltimore's win probability had dropped to 30%.

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

What implied volatility is consistent with pre-game market expectations?

To calculate the implied volatility of the Superbowl we substitute the pair $(\mu, p) = (-4, .353)$ into our definition and solve for σ_{IV} . We obtain

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

where we have used $\Phi^{-1}(p) = qnorm(0.353) = -0.377$. So on a volatility scale the 4 point advantage assessed for the 49ers is under a $\frac{1}{2}\sigma$ favorite. Another finding is that the inferred $\sigma = 10.6$ is historically low, as the typical volatility of an NFL game is approximately 14 points (see Stern, 1991). One possible explanation is that for a competitive game like this, one might expect a lower than usual volatility. In reality, 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 approach estimating 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 pre-game value). If so, then we can apply the formula derived earlier. The conditional mean for the

final outcome is $15 + 0.5 * (-4) = 13$ and the conditional volatility is $10.6\sqrt{1-t}$. These imply a probability of .96 for Baltimore to win the game.

A second estimate of the probability of winning given the half time lead can be obtained directly from the betting market. From the online betting market we also have traded contracts on TradeSports.com that yield a half time probability of $p_{\frac{1}{2}} = 0.90$. 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?

If we assume that p_t^{mkt} reflects all available information, we can use it to determine the market's assessment of implied volatility. For example, at half-time $t = \frac{1}{2}$ we would update

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

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 moneyline 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 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

$$p_{\frac{1}{2}} = \Phi\left(13/(10.6/\sqrt{2})\right) = 0.96$$

probability of winning versus the $p_{\frac{1}{2}}^{mkt} = 0.90$ 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 yields

$$\omega_{bet} = p_{\frac{1}{2}} - \frac{q_{\frac{1}{2}}}{O^{mkt}} = 0.96 - \frac{0.1}{1/9} = 0.60$$

where $p_{\frac{1}{2}}$ is our assessment of the probability of winning at half time (0.96), $q_{\frac{1}{2}} = 1 - p_{\frac{1}{2}}$ 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 high - it is mathematically correct since we believe we have a 96% chance of earning a 10% return but it is very sensitive to our estimated probability of winning. For this reason and others it is common 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 key injury. 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 $p_{\frac{1}{2}}^{mkt}$ 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 model for the evolution of sports scores and information from betting and prediction markets we are able to define the implied volatility of a sporting

event. Given the explosion of online betting and predictions markets there is an opportunity to learn from market-based information through the lens of a probabilistic model. There are two questions of interest. First, is the model reasonable? and, secondly, how informative is the market-information?

There are a number of features of the modeling assumptions that can be relaxed. One is the Markov nature of the probabilistic information. Is it really the case that the only relevant conditioning information is the current lead: $X(t) = l$ and not the past history? 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 would initially think. Of course, there are exceptions. For example, Jackson (1993) considers the non-Markovian nature of tennis games and the effect of non-independence on the assessment of rare event probabilities.

There have been many studies of the informational content of betting and prediction markets. See Snowberg et al (2012) for a recent discussion. 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 score in the game. For example, in SuperBowl XLVII the over-under line was 48 with the true score 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. Therefore, introducing payouts that are proportional to the magnitude of the points victory are more sensitive to the volatility of the

game and hence provide more information about volatility.

Our approach also sheds light on the relationship between the point spread and the moneyline odds. There is no one-to-one mapping as commonly thought. For example, it is not uncommon in college basketball to have two games both with a heavy favourite where the moneyline odds are the same but the point spread in one game is significantly higher than the other. Our explanation for this difference is a simple one—the market has an expectation that the volatility of the latter game is much higher.

We should remark that as with the Black-Scholes model for option prices, our definition of implied volatility doesn't rely on our model being true *per se*. It is merely a lens through which to communicate the bettors information about current expectations about the future course of the sporting event in the form of an implied volatility.

5 References

Avery, C. and Chevalier, J. (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.

Dixon, M.J. and S.G. Coles (1997). Modelling Association Football Scores and Inefficiencies in the Football Betting Market. *Appl. Statist.*, 46, 265-280.

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

Hull, J.C. (2005). *Options, Futures and Other Derivatives* (6th edition). Prentice Hall, New Jersey.

Jackson, D.A. (1993). Independent Trails are a Model for Disaster. *J. R. Statist. Soc. C*, 42(1), 211-220.

Jackson, D.A. (1994). Index Betting on Sports. *J. R. Statist. Soc. 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*, 917-926.
- Lee, A.J. (1997). Modeling scores in the Premier League: are Manchester United really the Best? *Chance*, 10, 15-19.
- Merton, R.C. (1976). Option Pricing when Underlying Stock Returns are Discontinuous. *Journal of Financial Economics*, 3, 125-144.
- Rosenfeld, J. W., J. I. Fisher, D. Alder and C. Morris (2010). Predicting Overtime with the Pythagorean Formula. *Journal of Quantitative Analysis in Sports*, 6(2), 1.
- Rosenfeld, J. W. (2012). An in-game win probability model of the NBA. *Thesis, Harvard University*.
- Snowberg, E., J. Wolfers and E. Zitzewitz (2012). Prediction Markets for Economic Forecasting. *Working Paper*.
- Speigelhalter, D.S. and Y-L. Ng (2009). One month to Go. *Significance*, 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.