The Valuation of Soccer Spread Bets

A.D. Fitt[*], C.J. Howls [†] & M. Kabelka[‡]

May 15, 2005

School of Mathematics, University of Southampton, Southampton SO17 1BJ, UK

Abstract

Simple statistical and probabilistic arguments are used to value the most commonly-traded online soccer spread bets. Such markets typically operate dynamically during the course of a match and accurate valuations must therefore reflect the changing state of the match. Both goals and corners are assumed to evolve as Poisson processes with constant means. Though many of the bets that are typically traded are relatively easy to value, some (including the "four flags" market) require more detailed analysis. Examples are given of the evolution of the spread during typical matches and theoretical predictions are shown to compare closely to spreads quoted by online bookmakers during some of the important matches of the EURO2004 tournament.

KEYWORDS: sports; gaming; recreation; finance; statistics

1 Spread betting

Spread betting on financial, sporting and other events as bizarre as the number of days that Saddam Hussein could avoid capture after his overthrow has, in spite of not being universally approved of (Wang and Ke, 2004), recently become extraordinarily popular. Viewed from a Financial Mathematics standpoint, a spread bet is simply a future (forward) contract. Share prices, at least by general consensus, follow a lognormal random walk. The growth rate and the volatility of a share or an index are well-documented

[*] **adf@maths.soton.ac.uk**
[†] **cjh@maths.soton.ac.uk**
[‡] **mk1602@soton.ac.uk**

and well-understood as concepts. In contrast, it is not immediately clear how runs scored by England in the first innings of a cricket test match may be modelled as a random process. Should we use historical data? Does it matter who England are playing, or what the weather and the wicket are doing? More obliquely, if an OB ("online bookmaker") wants to offer spread bets (of dubious taste) on how long Saddam will evade capture, how can s/he expect to find Excel-spreadsheet data on "desert capture rates for recently dethroned dictators"?

In this study we concentrate on the valuation of soccer spread bets. In the UK this rapidly-growing market is traded almost solely via OBs. We show how, using a combination of simple modelling and some elementary statistical theory, the dynamic spreads quoted by the bookmakers may be reproduced with a good deal of accuracy. All of the bets that we shall discuss can actually be traded, though for obvious reasons we will decline to name any specific OBs or URLs; if required, these may be rapidly identified using a simple internet search.

The key interest of soccer spread bets lies in their dynamic nature. As events take place during a game, bookmakers must alter the odds accurately to prevent arbitrage. Simultaneously the better must monitor his or her investment to ensure either that losses are controlled or winnings are protected. The valuations of the OBs may be tracked minute-by-minute as a match progresses. In this way the accuracy of the bookmakers' bet valuations (or our theory, depending on one's point of view) may be tested. Our main aims are to understand how rapidly valuations can change, to determine whether or not OBs are modelling goals and corners as Poisson processes, and to be able to identify circumstances where OBs are quoting bet valuations at variance with the theory.

1.1 The mechanics and rules of spread betting

Before turning to spread bet valuation *per se*, we give brief details of the basic mechanics of spread bets. Normally, the betting process begins when one registers with an OB. A sum of money is sent to the OB who opens an online account (some pay interest) which bankrolls all transactions.

Unlike a standard "fixed odds" bet, spread bets are volatile and may involve large (sometimes potentially unbounded) wins or losses. Perhaps the simplest traded soccer bet concerns the time (in minutes) of the first goal in a match. Before the game starts the OB will quote a spread of (for example) 36-39 minutes. Bets are reckoned in standard wager amounts w per minute (typically $w = £10$). "Buying the spread" at 39 minutes assumes that one takes the view that the first goal will occur later than 39 minutes into the game. If the first goal is actually scored after N minutes then the (better's) payoff P for this bet is given by $P = w(N-39)$ (and may be either positive or negative). Conversely, "selling the spread" reflects a view that n is likely to be less than 36: the payoff P for this bet is $P = w(36 - N)$.

A few other points are worth making: operationally, both wins and losses are credited to one's online account. Entering each bet is free, for the spread gives the OB their profit (if the outcome lies within the spread, then $P < 0$ for both the buyers and the sellers). At any time, a better may "close out" and guarantee a fixed profit or loss simply by taking the opposite position to their current bet. The betting transaction then ends as the positions simply cancel each other out. If, during a bet, an account incurs losses that reduce it to zero, the OB will automatically close all positions and suspend the account until more funds are deposited.

For theoretical purposes the size of the spread is irrelevant since it is fixed in advance (according to market conditions) by the bookmakers to control their profits. To value a bet, it is therefore enough to determine the "centrespread" C, which, because of the dynamic nature of spread bets as they are traded while the match is in progress, is in general a function of time.

1.2 Common soccer markets

In this study we will attempt to value some of the most commonly-traded soccer spread bets (most bets may also be traded for individual teams, where analogous valuations easily follow). These include:

- **Total goals/corners:** The total number of goals or corners scored or taken in any given match by both teams.
- **Goal/corner supremacy:** A given team's "dominance" over their opposition, measured as the difference between the two team's goals/corners that are scored/taken. (May be negative.)
- **Multicorners:** The product of the total number of corners taken in each half of the match.
- **Crosscorners:** The product of the total home team corners and the total away team corners.
- **Time of nth goal:** Self-explanatory, but is settled at 90 if the nth goal is not scored (added time being irrelevant).

- **Time of final goal:** Added time is irrelevant and if no goal is scored the bet is settled at 0.
- **Total goal minutes:** The sum of the minutes when each goal in the game is scored. (A 1-1 draw with goals in the 34th and 80th minute would give 114 goal minutes.)
- **Four flags:** The time in minutes when corners have been taken from all four corner positions (settled at 90 if a flag remains unused). Note that this bet is not normally traded during the game: we speculate that this is because OBs cannot value it.

Other types of bet may also be offered, and these may be valued using similar methodology to that presented below. Visits to the web site of the most popular OBs will reveal that bets appear to go "in and out of fashion" and may sporadically disappear as offered products, only to reappear a short time later.

2 Spread bet valuation

To value the common soccer spread markets listed in Section 1.2, we will follow many previous studies (Dyte and Clarke, 2000), (Maher, 1982) and assume that both the scoring of goals and the occurrence of corners are Poisson processes with constant mean. It is a simple matter to investigate the validity of the this assumption using real data.

Figure 1 shows the distribution of all goals scored in the 380 English Premiership matches for the season 2002-3. The fit to the Poisson model is obviously close. Similar results may be also be observed for corners, other seasons, other countries and other competitions (though see Section 4). There is some evidence that under certain circumstances the Poisson mean μ changes during the match. For the purposes of this study however we will investigate neither methods of estimating μ nor time-dependent formulae for μ. All matches will be assumed to last for a duration of 1 time unit, which includes any added time that may occur: all matches commence at $t = 0$. If X is a random variable then $E_t(X)$ denotes the expected value of X at time t, $N(a, b)$ denotes the number of occurrences of a given event during the time period (a, b) and Poisson means for the whole match are denoted by μ so that

$$P(N(a, b) = n) = \frac{(\mu(b - a))^n e^{-\mu(b-a)}}{n!}.$$

2.1 Total goals

Valuation of the total goals spread bet is simple. Denoting the total number of goals by N, we know that at $t = 0$, $C(0) = E_0(N) = \mu$ where here and henceforth we denote the value of the centre of the spread at time t by $C(t)$. If exactly r goals have been scored in the period $(0, t)$, then

$$C(t) = E_t(N \mid N(0,t) = r) = \sum_{n=0}^{\infty}(r+n)\frac{(\mu(1-t))^n e^{-\mu(1-t)}}{n!} = r + \mu(1-t).$$

Thought of as a function of μ, r and t we therefore have, for $\mu > 0$, $r \in \mathbb{N}$ and $0 \leq t \leq 1$

$$C(\mu, r, t) = r + \mu(1-t). \tag{1}$$

The centrespread therefore takes the value μ before the game begins and is thereafter a straight line with slope $-\mu$ which jumps by an amount $+1$ whenever a goal is scored, as shown in the example considered in figure 2. This simple result may be established in many other ways: for example, at time t the expected number of goals in the remainder of the match is simply $\mu(1-t)$, giving the expected number of goals, and therefore $C(t)$ as $\mu(1-t) + r$. Note that the very simplicity of this result is appealing insomuch as it allows the Poisson means being used by a given OB to be easily "backed out" before a match starts so that other, more complicated bets may be accurately valued.

2.2 Goal supremacy

We now define X as the goal supremacy of the home team (team A, say) against the away team (team B). If A and B score goals with Poisson means μ_A and μ_B respectively, we have $C(0) = \mu_A - \mu_B$. The centrespread for goal supremacy is evidently the mean of a difference of Poisson means, so that

$$C(\mu_A, \mu_B, r, t) = r + (1-t)(\mu_A - \mu_B). \tag{2}$$

The quantity C therefore takes the value $\mu_A - \mu_B$ before the match begins and is thereafter a straight line with gradient $\mu_A - \mu_B$ and vertical jumps of $+1$ whenever team A scores and -1 whenever team B scores.

The formulae (1) and (2) also apply respectively to total corners and corner supremacy when the appropriate Poisson means are used. If desired, (2) may also be established directly as in Section 2.1, though it is necessary to use some elementary properties of Bessel functions to establish the result.

2.3 Multicorners

To value the multicorner spread bet, we let X denote the product of the first and second half total corner counts and use $N(a, b)$ to denote the total number of corners taken during the time period (a, b). Evidently $X = N(0, \frac{1}{2})N(\frac{1}{2}, 1)$ and $E_0(X) = E(N(0, \frac{1}{2}))E(N(\frac{1}{2}, 1))$. Denoting the Poisson mean of the total number of match corners by μ, we have $E(N(0, \frac{1}{2})) = E(N(\frac{1}{2}, 1)) = \frac{\mu}{2}$, whence $E_0(X) = \frac{\mu^2}{4}$.

For $0 \leq t \leq \frac{1}{2}$ we have, assuming that r corners have taken place during $[0, t]$,

$$E(X \mid N(0,t) = r) = \left[\sum_{n=0}^{\infty}(r+n)P(N(t,\frac{1}{2}) = n)\right]\left[\sum_{n=0}^{\infty} n\ P(N(\frac{1}{2},1) = n)\right]$$

$$= \left[\sum_{n=0}^{\infty}(r+n)\frac{(\mu(\frac{1}{2}-t))^n e^{-\mu(\frac{1}{2}-t)}}{n!}\right] E(N(\frac{1}{2},1)) = \left(r + \mu\left(\frac{1}{2} - t\right)\right)\left(\frac{\mu}{2}\right)$$

whence

$$C(\mu, t) = \frac{\mu^2}{4} - \frac{\mu^2}{2}t + \frac{\mu}{2}r \qquad \text{for} \quad 0 \leq t \leq \frac{1}{2}.$$

Thus $C = \mu^2/4$ before the match commences and, during the first half, is a straight line with gradient $-\mu^2/2$ and vertical jumps of $+\frac{\mu}{2}$ whenever a corner is taken.

During the second half, when $\frac{1}{2} < t \leq 1$, we already know the number of corners R_1, say, taken in the first half. Thence $C(t) = R_1(r - R_1 + \mu(1-t))$, and C is a straight line of slope $-R_1\mu$ with vertical jumps of $+R_1$ whenever a corner is taken. Thus for multicorners

$$C(\mu, R_1, r, t) = H(\frac{1}{2} - t)\left(\frac{\mu^2}{4} - \frac{\mu^2}{2}t + \frac{\mu}{2}r\right) + H(t - \frac{1}{2})R_1(r - R_1 + \mu(1-t)) \tag{3}$$

where H denotes the unit Heaviside step function. A typical multicorners valuation profile is shown for a particular case in figure 3. As might be expected, at half time ($t = 1/2$), the value of the bet is continuous but its slope is discontinuous. Figure 4 shows a comparison between (3) and the dynamically quoted centrespread of a well-known OB for the Portugal multicorners during the crucial Portugal vs. Holland game in the Euro2004 tournament: the agreement between (3) and the OB's quoted centrespread is clearly very marked.

2.4 Crosscorners

The crosscorner spread bet is now easy to value. We have simply

$$C(\mu_A, \mu_B, r_A, r_B, t) = [r_A + \mu_A(1-t)][r_B + \mu_B(1-t)] \tag{4}$$

where r_A and r_B are the respective number of corners so far taken by each of the two teams, which are assumed to win corners with Poisson means μ_A and μ_B. Here the centrespread takes the value $\mu_A\mu_B$ at the start of the match and thereafter is a sequence of parabolas, with vertical jumps of $r_B + \mu_B(1-t)$ whenever team A takes a corner and $r_A + \mu_A(1-t)$ whenever team B does. Comparisons between (4) and the centrespreads set by OBs during important matches (not shown) revealed good agreement. The jumps depend on the time elapsed and the numbers of corners so far taken. From a betting point of view it is important to understand that if the team(s) have already taken many corners, the bet is a volatile and risky one. If, however few corners have previously been taken and the end of the match is nearing, the centrespread is likely to remain relatively constant.

2.5 Time of nth goal

The centrespread for bets that involve times of events may be valued by considering the Poisson Cumulative Distribution Function (CDF). At time t, we denote the time of the n^{th} goal by t_n, and consider the probability $F_n(t,x)$ that at least n goals will occur during the time period (t,x), given that no goals occurred during $(0,t)$. We have, for $0 \leq t < x < 1$,

$$F_n(t,x) = P(t \leq t_n \leq x) = 1 - P(t_n > x \mid N(0,t) = 0) = 1 - P(N(t,x) < n)$$
$$= 1 - \sum_{k=0}^{n-1} P(N(t,x) = k) = 1 - \sum_{k=0}^{n-1} \frac{(\mu(x-t))^k e^{-\mu(x-t)}}{k!}$$

while $P(t_n \leq x) = 1$ for $x \geq 1$. (We recall that if fewer than n goals are scored the time of the n^{th} goal is assumed to be $t = 1$.) The PDF f_n of t_n may now be found by differentiation of the CDF, yielding, for $0 \leq x < 1$,

$$f_n(t,x) = -\sum_{k=0}^{n-1} \frac{\partial}{\partial x} \frac{(\mu(x-t))^k e^{-\mu(x-t)}}{k!} = \frac{\mu^n (x-t)^{n-1}}{(n-1)!} e^{-\mu(x-t)} \tag{5}$$

which is a Gamma distribution, parameter-shifted by an amount t. The probability of the market settling at $t = 1$ is simply the probability that fewer than n goals will be scored in time $(t,1)$. Thus $P(t_n = 1) = 1 - F_n(t,1)$.

The expected time of the n^{th} goal at time t may now be found by multiplying $f_n(t,x)$ by x and integrating from $x=t$ to $x=1$, giving

$$C(n,\mu,t) = \frac{\mu^n e^{\mu t}}{(n-1)!}\int_t^1 x(x-t)^{n-1}e^{-\mu x}dx + e^{-\mu(1-t)}\sum_{k=0}^{n-1}\frac{(\mu(1-t))^k}{k!}. \quad (6)$$

The expressions involved (which may be written in terms of incomplete gamma functions) become rapidly more unwieldy as n increases: even at the start of the match $C(n,\mu,0)$ may be succinctly expressed only in terms of hypergeometric functions. For the commonly-traded time of first and second goal markets, however, we find that

$$C(1,\mu,t) = E_t(t_1) = t + \frac{1}{\mu} - \frac{e^{-\mu(1-t)}}{\mu},$$

$$C(2,\mu,t) = E_t(t_2) = t + \frac{2}{\mu} + \frac{e^{-\mu(1-t)}}{\mu}(\mu(t-1)-2).$$

Note also that asymptotic results are easy to derive. Some manipulation of (6) shows that in fact

$$C(n,\mu,t) = 1 - \frac{\mu^n e^{\mu t}}{(n-1)!}\int_t^1 (1-x)(x-t)^{n-1}e^{-\mu x}\,dx. \quad (7)$$

For the asymptotic limit $n \gg 1$ two integrations by parts now reveal that

$$C \sim 1 - \frac{(\mu e)^n e^{-\mu(1-t)}(1-t)^{n+1}n^{-n-3/2}}{\sqrt{2\pi}} \quad (n\to\infty)$$

so that for large n the time of the nth goal is very close to unity. Close to the end of the match, we set $t = 1-\epsilon$ where $\epsilon \ll 1$. It is then easy to show from (7) that

$$C \sim 1 - \frac{\mu^n \epsilon^{n+1}}{(n+1)!} \quad (t\sim 1)$$

so that again the centrespread is very close to unity.

Typical valuation curves for the centrespread are shown for $n = 1,2,3,4$ in figure 5. Note that (6) implicitly assumes that no goal was scored in time $(0,t)$. If it should happen that a goal is scored, then C jumps down to the "time of $(n-1)^{th}$ goal" curve. In the general case where r goals have been scored, we simply locate $E_t(t_{n-r})$, which follows the curve for the $(n-r)^{th}$ goal. An example of the valuation of the time of the 2nd goal is given in figure 6. A comparison between the observed spreads of an OB for the time of England's 2nd goal during the England vs. Croatia Euro2004 match is shown in figure 7. The close agreement between the quoted spreads and the theoretical predictions (symbols) is clear: in this case the market closed when England scored their 2nd goal in the 47th minute.

2.5.1 Discrete valuation for time of nth goal

Thus far, we have assumed that this bet may be settled at any value between 0 and 1. When traded in practice, however, the valuation is calculated using only discrete values, for if a goal is scored after (say) $t = 23.3$ minutes the bet is settled at 24. In order to recalculate $E_t(t_1)$ (other values of n may be similarly treated) taking this discretisation into account, we assume that the bet is settled only at values $t = \frac{n}{90}$, $n = 1, ..., 90$. Suppose, therefore, that at time $0 \leq t \leq 1$ no goals have yet been scored and $\lceil 90t \rceil = m$ where here $\lceil x \rceil$ denotes the smallest integer greater than x (the "ceiling" function). Then writing $P(t_D = n)$ to denote the probability that the time of first goal bet will be discretely settled at minute n we have, using the PDF (5),

$$P(t_D = n) = \int_{(n-1)/90}^{n/90} \mu e^{-\mu(x-t)} dx = e^{\frac{\mu}{90}(1-n+90t)} - e^{\frac{\mu}{90}(-n+90t)}$$

for $n = m + 1, m + 2, ..., 90$, and

$$P(t_D = n) = \int_{t}^{m/90} \mu e^{-\mu(x-t)} dx = 1 - e^{\frac{\mu}{90}(-m+90t)}$$

for $n = m$. The expected value, and thus the centrespread, may now be determined by integration in the usual way. We find that

$$C(1, \mu, t) = E_t(t_1) = e^{-\mu(1-t)} + \frac{m}{90}\left(1 - e^{\frac{\mu}{90}(-m+90t)}\right) +$$

$$\sum_{n=m+1}^{90} \frac{n}{90}\left(e^{\frac{\mu}{90}(1-n+90t)} - e^{\frac{\mu}{90}(-n+90t)}\right).$$

Similar calculations may be carried out for the time of the n^{th} goal, but computations show that the difference between continuous and discrete valuation is probably too small to be of significant financial use to the average "punter".

The inclusion of the influence of added ("injury") time is more important. According to market rules, a goal scored in added time at the end of either half is settled at 45 or 90 minutes. Settlement at either of these times should therefore be slightly more probable. Assuming that we know the expected added time for each half, we assume that periods I_1 and I_2 respectively are added at the end of each half, so that the time period $(0,1)$ must now correspond to $T = 90 + I_1 + I_2$ minutes. Using similar reasoning to above, we have, for t during the first half when no goals have yet been scored and $\lceil 90t \rceil = m$, (other cases may be treated similarly)

$$C(1,\mu,0) = E_t(t_1) = m(1 - e^{-\frac{\mu}{T}(m-tT)}) + 45(e^{-\frac{\mu}{T}(44-tT)} - e^{-\frac{\mu}{T}(45+I_1-tT)}) +$$

$$\left(\sum_{n=m+1}^{44} n\left(e^{-\frac{\mu}{T}(n-1-tT)} - e^{-\frac{\mu}{T}(n-tT)}\right)\right) +$$

$$\left(\sum_{n=46}^{89} n\left(e^{-\frac{\mu}{T}(n+I_1-1-tT)} - e^{-\frac{\mu}{T}(n+I_1-tT)}\right)\right) + 90e^{-\frac{\mu}{T}(89+I_1-tT)} \qquad (8)$$

where values have been expressed in real time minutes, i.e. in a range $0-90$. Evaluation shows that, when the added time is relatively small, its effect is negligible. However, the data contained in table 1 where the expected time of the first goal is calculated at the start of a match using a variety of different assumptions, shows that the centrespread may change by a non-trivial amount if (for example) the total amount of added time is 12 minutes. Such a circumstance is uncommon, but by no means rare.

2.6 Time of the last goal

Though we have valued the time of the n^{th} goal, we do not know *a priori* how many goals will be scored in total. To value the time t_ℓ of the last goal we assume that no goal has been scored during the time period $(0,t)$ and that the bet can be settled at any value between 0 and 1 (recall that if no goals are scored in the entire match the bet is settled at $t = 0$). We have

$$F(t,x) = P(t_\ell = 0) + P(t \le t_\ell \le x) = 1 - P(t_\ell > x)$$

$$= 1 - P(N(x,1) \ge 1) = P(N(x,1) = 0) = e^{-\mu(1-x)},$$

which is independent of the current time t. For $t \le x \le 1$, differentiation gives $f(t,x) = \mu e^{-\mu(1-x)}$ and $P(t_\ell = 0) = P(N(t,1) = 0) = e^{-\mu(1-t)}$. For this bet, C is therefore determined by the expected value

$$C(\mu,t) = E_t(t_\ell) = 0e^{-\mu(1-t)} + \int_t^1 x\,\mu e^{-\mu(1-x)} dx$$

$$= 1 - \frac{1}{\mu} + e^{-\mu(1-t)}\left(\frac{1}{\mu} - t\right).$$

We note that $E_0(t_1) + E_0(t_\ell) = 1$, so that, at the start of the match the centrespreads for the times in minutes of the first and last goals should sum to 90. Put another way, in the absence of extra time, the time of the last goal could be derived by "running the match backwards".

At the start of the match we have $C = 1 - (1 - e^{-\mu})/\mu$, but how should the centrespread for this bet be adjusted when a goal is scored? Assume that time t has elapsed in the match and that the most recent goal was scored at time $t_r < t$ (where $t_r = 0$ if no goals have yet been scored). For $t \le x \le 1$, we have $f(t, x) = \mu e^{-\mu(1-x)}$ and $P(t_\ell = t_r) = P(N(t, 1) = 0) = e^{-\mu(1-t)}$. Thus, for $0 \le t_r \le t < 1$

$$C(\mu, t_r, t) = E_t(t_\ell) = t_r e^{-\mu(1-t)} + \int_t^1 x\mu e^{-\mu(1-x)}\, dx$$

$$= 1 - \frac{1}{\mu} + e^{-\mu(1-t)}\left(\frac{1}{\mu} - t + t_r\right),$$

so that whenever a goal is scored the centrespread jumps by an amount $e^{-\mu(1-t)}(t_r^{new} - t_r^{old})$. Typical valuation curves for this market are shown in figure 8: the bet is a volatile one if there is an extended goalless period during a match.

Similar remarks as in Section 2.5.1 apply regarding the difference between discrete and continuous valuations of this bet. The probability that the final goal will be scored in the m^{th} minute, so that the bet is settled at $t_D = m/90$ may be calculated by integrating $f(t, x)$. Assuming once again that a time t has elapsed since the start of the match where $\lceil 90t \rceil = m$ and no goals have yet been scored, we find that

$$C(\mu, t) = E_t(t_\ell) = \frac{m}{90}\left(e^{\frac{\mu}{90}(m-90)} - e^{-\mu(1-t)}\right) +$$

$$\sum_{n=m+1}^{90} \frac{n}{90}\left(e^{\frac{\mu}{90}(n-90)} - e^{\frac{\mu}{90}(n-91)}\right).$$

2.7 Total goal minutes

To value this bet, it is easiest to define X as the sum of the goal times and consider the quantity $E(X \mid N(a, b) = n)$, namely the expected value of the total goal minutes during (a, b) given that n goals were scored during the time period (a, b). If n is zero, then evidently X will always be zero. If

$n = 1$ then since we have assumed that goals occur via a time-homogeneous Poisson process, we would expect the goal to occur at the midpoint of the interval, so that $E(X \mid N(a,b) = 1) = (a+b)/2$. If $n = 2$ then by similar reasoning $E(X \mid N(a,b) = 2) = 2(a+b)/2$, the first goal being scored at $t = (2a+b)/3$ and the second at $t = (a+2b)/3$. Thence in general $E(X \mid N(a,b) = n) = \sum_{k=1}^{n} a + k(b-a)/(n+1) = n(a+b)/2$.

To calculate the expected value of X at any time t during a match, assuming that no goals have been scored during the time $(0,t)$, we have, by summing expectations,

$$E_t(X) = \sum_{n=0}^{\infty} E(X \mid N(t,1) = n) P(N(t,1) = n)$$

$$= \sum_{n=0}^{\infty} \frac{n(t+1)(\mu(1-t))^n}{2n!} e^{-\mu(1-t)}$$

so that

$$C(\mu,t) = \frac{\mu}{2}\left(1-t^2\right).$$

At the start of the match $C = \mu/2$, a fact obvious from Poisson equipartition of events. Each time a goal is scored, the centrespread should increase by the goal time. Thence for this bet

$$C(\mu,r,t) = E_t(X) = \frac{\mu}{2}\left(1-t^2\right) + r_T \tag{9}$$

where r_T is the sum of the goal minutes of all goals thus far scored.

Typical valuation curves for this bet are shown in figure 9. The total goal minutes market is regarded by OBs as one of the most volatile soccer spread bets since goals near to the end of the match may change the settlement of the bet by large amounts. Figure 10 shows a comparison between the total goal minute spreads quoted by an OB and (9) for the England vs. Croatia match in the Euro2004 tournament. Once again, the close agreement between the theory and the observed spreads is clear.

2.8 Four flags

Many of the centrespread valuations presented thus far have relied on relatively simple arguments. The "four flags" bet, however, turns out to be somewhat more involved. To value it, we assume that, in addition to the fact that each team's corners are Poisson processes with means μ_A and μ_B, each corner awarded is equally likely to be taken from the right or the left corner flag. We denote the corner flags from where team A takes corners in the first half of the match by A_1 and A_2. Corners at flags A_1 and A_2 are therefore Poisson processes with means $\mu_A/2$ and $\mu_B/2$ in the first and second halves respectively. Similar notation is used for flags B_1 and B_2. We further define $F_{A_1}(t,x)$ as the CDF that gives the probability that at least one corner is taken from flag A_1 in the time period (t,x): this is equivalent to $P(t \leq t_{1A1} \leq x)$, where t_{1A1} is the time of the first corner taken from flag A_1. The functions $F_{A_2}(t,x)$, $F_{B_1}(t,x)$ and $F_{B_2}(t,x)$ are similarly defined.

We may now evaluate $F_{A_1}(t,x)$ in several parts, depending on the values of t and x. Temporarily, we write $N(a,b)$ to denote the number of corners taken from flag A_1 in time (a,b). We then have:

$$(0 \leq t \leq x \leq 1/2): \quad F_{A_1}(t,x) = P(N(t,x) \geq 1) = 1 - e^{\frac{-\mu_A}{2}(x-t)}$$

$$(0 \leq t \leq 1/2 \leq x \leq 1): F_{A_1}(t,x) = P\left(t \leq t_{1A1} \leq \frac{1}{2}\right) + P\left(\frac{1}{2} < t_{1A1} \leq x\right)$$

$$= P(N(t,\frac{1}{2}) \geq 1) + P(N(t,\frac{1}{2}) = 0)P(N(\frac{1}{2},x) \geq 1 \mid N(t,\frac{1}{2}) = 0)$$

$$= 1 - P(N(t,\frac{1}{2}) = 0) + P(N(t,\frac{1}{2}) = 0)P(N(\frac{1}{2},x) \geq 1)$$

$$= 1 - e^{\frac{-\mu_A}{2}(\frac{1}{2}-t) - \frac{\mu_B}{2}(x-\frac{1}{2})}$$

(since, after half time, it is team B that takes corners from flag A_1)

$$(1/2 \leq t \leq x \leq 1): \quad F_{A_1}(t,x) = P(N(t,x) \geq 1) = 1 - e^{\frac{-\mu_B}{2}(x-t)}.$$

Thus

$$F_{A_1}(t,x) = H\left(\frac{1}{2}-t\right) H\left(\frac{1}{2}-x\right)\left(1 - e^{\frac{-\mu_A}{2}(x-t)}\right) + H\left(\frac{1}{2}-t\right) H\left(x-\frac{1}{2}\right)\left(1 - e^{\frac{-\mu_A}{2}(\frac{1}{2}-t) - \frac{\mu_B}{2}(x-\frac{1}{2})}\right) + H\left(t-\frac{1}{2}\right)\left(1 - e^{\frac{-\mu_B}{2}(x-t)}\right).$$

Since corners at flags A_1 and A_2 have the same probability of being taken in any given time interval, $F_{A_1}(t,x) = F_{A_2}(t,x)$: we denote this quantity by $F_A(t,x)$. Similarly, $F_{B_1}(t,x) = F_{B_2}(t,x) = F_B(t,x)$, where $F_B(t,x)$ has the same form as $F_A(t,x)$, but with all occurrences of μ_A replaced by μ_B and *vice versa.*

Now define X as the time when the "fourth flag" is first used. At time t, we assume that no corners have so far taken place. Then (note that, by definition, $F(t,t)=0$),

$$\begin{aligned} F(t,x) &= P(X \le x) \\ &= P(t \le t_{1A1} \le x)P(t \le t_{1A2} \le x)P(t \le t_{1B1} \le x)P(t \le t_{1B2} \le x) \\ &= F_{A_1}(t,x)F_{A_2}(t,x)F_{B_1}(t,x)F_{B_2}(t,x) \\ &= F_A^2(t,x)F_B^2(t,x). \end{aligned}$$

Since the bet is settled at $t=1$ if the fourth flag is never used, the expected value of X is given by

$$E_t(X) = \int_t^1 x\frac{\partial}{\partial x}F(t,x)\;dx + 1 - F(t,1) = 1 - \int_t^1 F(t,x)\;dx. \tag{10}$$

At the start of the match, the centrespread therefore takes the value

$$C(\mu_A,\mu_B,0) = \frac{Q}{psc} \tag{11}$$

where

$$Q = a^4b^4pc - 12a^3b^3ps^2 + a^2(b^2c(\mu_A^2+10p+\mu_B^2)+4bpd-dc)+$$

$$a(-4b^2dp+4sc(q-sb))+(b(b-4)dc+3(-3p+2s^2)(p+s^2))$$

and $c=(2\mu_A+\mu_B)(2\mu_B+\mu_A)$, $d=\mu_A^2-\mu_B^2$, $s=\mu_A+\mu_B$, $p=\mu_A\mu_B$, $q=\mu_A-\mu_B$, $a=e^{-\mu_A/4}$ and $b=e^{-\mu_B/4}$. Figure 11 shows how the start-of-match centrespread given by (11) changes for various values of μ_A and μ_B. The function decreases with any increase in either of the parameters, and it is also clear that, as might be expected, $E_0(X)\to 1$ as μ_A, $\mu_B\to 0$, $E_0(X)\to 0$ as μ_A, $\mu_B\to\infty$ and $E_0(X)\to\frac{1}{2}$ as $\mu_A\to\infty$, $\mu_B\to 0$ or $\mu_A\to 0$, $\mu_B\to\infty$.

Though (10) values the four flags bet at any time t provided no corner has yet been awarded, our main aim is dynamic valuation of this bet. How, therefore, should (10) be modified if and when corners are awarded? Once a corner is taken from flag A_1, its further use is irrelevant, for all that now matters is when the remaining three flags are used. Once A_1 is used, therefore, $F(t,x)$ should become a multiple of the three CDF's of the remaining flags. Accordingly we define four new indicator functions r_{A_1}, r_{A_2}, r_{B_1} and

r_{B_2}. Each is unity when the match starts and each becomes zero when a corner is first taken from the corresponding flag. Thence

$$F(t,x) = F_A^{r_{A_1}+r_{A_2}}(t,x) F_B^{r_{B_1}+r_{B_2}}(t,x)$$

and $E_t(X)$ may easily be calculated as in (10).

As the match progresses, there are now eight different possibilities for the evolution of $C(t)$, depending on the exponents of $F_A(t,x)$ and $F_B(t,x)$ in the definition of $F(t,x)$.

For convenience, let $r_{A_1} + r_{A_2} = r_A$ and $r_{B_1} + r_{B_2} = r_B$. Each of these exponents takes the value 0, 1 or 2, depending on how many corners have already been taken from flags on the corresponding side of the pitch. Excluding the case $r_A = r_B = 0$, where all four flags have been used and the bet is already settled, this gives eight possible combinations.

Samples of these curves are shown in figure 12, where the two corner means have been chosen to be significantly different. It is probably easiest to regard the curves in this figure as "railtracks" along which the valuation runs, the choice of track being dependent on the order in which the corner flags are used. Suppose, for example, that $t \geq \frac{1}{2}$ and three flags have already been used. Then C follows the same curve as the "time of first goal" market but with parameter $\mu_A/2$ or $\mu_B/2$, depending on which team will take corners from the remaining flag.

Note that if $\mu_A \sim \mu_B$, then the curve for $E_t(X)$ is similar in shape to the curve for the "time of n^{th} goal" market: an example of such a valuation curve is shown in figure 13 where $\mu_A = 5.2$ and $\mu_B = 6.1$. However, if mismatched teams meet so that one value of μ is much in excess of the other, then the valuation curve can be volatile and very different to that of any other market so far examined. In figure 14 the large difference between μ_A and μ_B leads to an exotically-shaped valuation curve which is part concave, part convex and has a marked slope discontinuity at half time. Evidently an intimate understanding of this valuation is required if the bet is to be traded dynamically.

3 Risk and reward

Having calculated the "fair value" for the commonest football spread bets, some assessment of the associated risk is apposite. In each case that we have already examined, the variance may be calculated in the obvious way. For example, for the simple total goals bet, we have seen that

$$C(\mu, r, t) = r + \mu(1-t)$$

where r is the number of goals that have already been scored at time t. Since the process is of Poisson type, the variance (i.e. the square of the risk) is simply $\mu(1-t)$ (the mean number of goals from time t to the end of the match). The centrespread is essentially calculated by determining where the average return on the bet is zero, thus from a risk/reward perspective one might argue that one should bet whenever the risk is low. The logical conclusion of this argument tells us that bets should be made very close to the end of the match, but unfortunately the existence of the spread means that, in reality, one is virtually certain to lose money in these circumstances as the fair value of the bet always lies within the spread.

Can calculating the variance tell us anything of practical value for more complicated bets? In contrast to the total goals bet, for multicorners the variance depends on r, the number of events (corners) that have already occurred. The variance is therefore a sequence of monotone decreasing functions connected by upward jumps that take place whenever events occur. This simply emphasises that the multicorner bet becomes more risky as more corners occur.

For online investors who wish to choose their level of risk, knowing the variance of a given bet can be more helpful. For example, the variance of the total goal minutes bet may be shown to be

$$\sigma^2 = \frac{\mu}{3}(1-t^3).$$

A risk averse investor should evidently conclude that, since the risk of this bet decreases relatively slowly until half time, second half wagers may better suit their betting philosophy.

For the time of the last goal the risk increases until it achieves a local maximum, after which it rapidly decreases to zero. When a goal is scored, the risk always decreases (by jumping down) . A rational better should therefore bet either at the start of the match or just after a goal is scored. In contrast, the risk for the time of the nth goal may jump up *or* down (dependent on the goal history) after a goal is scored.

In general, spread betters may use the risk curve in two ways: the investor who trades according to what s/he regards as an "acceptable level of risk" can use the risk curve to time their bet optimally. Alternatively, the risk curve can be used to determine where the risk changes most rapidly, since a mistimed bet in these regions (which often occur near to the end of a match) may prove to be very costly.

4 Conclusions

Simple probability theory has been used to value the most commonly-traded soccer spread bets. Although the methodology involved is relatively straightforward, to our knowledge explicit formulae for soccer spread bet valuations of this type have not appeared in the scientific literature before. Comparisons with data observed during some recent important matches show that our valuations are consistently close to those used by OBs.

A number of possibilities present themselves for further investigation: for example, it would also be possible to value all of the bets considered above by direct simulation. This would provide a simple means of testing any strategies that may evolve as a result of our valuations. It would also be valuable to determine the accuracy of the Poisson distribution assumptions that have been made for both goals and corners. Some authors have suggested that the Poisson distribution may not be the most appropriate one to use for these events, and others have questioned the validity of assuming that the Poisson mean is constant. For example, evidence has previously been presented (Bennett, 1998), (Greenhough et al., 2002) to suggest that the occurrence of goals may depend on certain conditions and events that may take place during the game (for example the dismissal of a player (Ridder et al., 1994), and particularly on the current score (Dixon, 1998)). If (for example) it really is true that scoring rates tend to increase as a game progresses, then this would increase the centrespread for many of the commonest bets such as total goals, goal minutes and time of last goal. If (as comparisons suggest) the OBs assume that the associated Poisson parameters are constant, then this misvaluation would clearly give the online better an "edge" over OBs. Similarly, if a team substituted a particularly active right winger at half time with a highly-rated left winger, the "four flags" centrespread would also be increased. The reader keen to experiment will find a wealth of both soccer and betting statistics at www.soccerbase.com.

For practical purposes, one of the most important issues for online betters is the matter of estimating the Poisson means and other data that are used in the predicted centrespreads. The fact that most of the data is readily available means that one realistic strategy for parameter estimation is to simply use historical values. It is likely however that a superior plan would be to infer the values required from other spreads being quoted. This could be done either by examining the valuations of different OBs for the same bet or by using centrespread values quoted by a single OB for different (but related) markets. The latter approach was used to produce the comparison figures 4, 7 and 10: the parallels between the choice of whether to use historical or implied volatility for the Black-Scholes valuation of financial derivative products are obvious. Clearly, if inferred (rather than historical) values are to be used the availability of closed-form centrespread formulae (such as those derived in this paper) is crucial.

Finally, it is worth pointing out that theoretically perfect valuation may not always be desirable for either the OB or the better. The matter of optimising one's overall betting strategy and the search for "arbitrage" has been addressed in previous studies (Haigh, 2000), (Finkelstein and Whitely, 1981), (Arends, 2002) as have the issues of "sentimental traders" and "false experts" (Avery and Chevalier, 1999) and may lead to non-optimal behaviour. In addition, other aspects of sporting spread bet markets have previously been considered (Jackson, 1994). Also, as equity market makers sometimes "shake the tree" to increase interest in a particular financial product, so OBs may adjust the spread non-optimally to encourage investment. As an OB sets the spread they must also consider how much has already been invested in buys and sells. If the disparity between these investments is great, then it may make good business sense to adopt a valuation that is at variance with theory.

Acknowledgement

Martin Kabelka gratefully acknowledges the support of the Nuffield Foundation.

References

1 Arends, B (2002). Spread Betting: A Football Fan's Guide. Time Warner Books: London.

2 Avery C and Chevalier J (1999). Identifying sentiment from price paths: The case of football betting. J Bus 72: 493–521.

3 Bennett J (1998). Statistics in Sport. Arnold: London.

4 Dixon M J (1998). A birth process model for association football matches. J Roy Stat Soc D 47: 523–528.

5 Dyte D and Clarke S R (2000). A ratings based Poisson model for World Cup soccer simulation. J Oper Res Soc 51: 993–998.

6 Finkelstein, M and Whitely, R. (1981). Optimal strategies for repeated games. Adv Appl Probab 13: 415–428.

7 Greenhough J, Birch P C Chapman S C and Rowlands G (2002). Football goal distributions and extremal statistics. Physica A 316: 615–624.

8 Haigh J (2000). The Kelly criterion and bet comparisons in spread betting. J Roy Stat Soc D 49: 531–539.

9 Jackson DA (1994). Index betting on sports. The Statistician 43: 309–315.

10 Maher M J (1982). Modelling association football scores. Statist Neerland 36: 109–118.

11 Ridder G, Cramer J S and Hopstaken P (1994). Estimating the effect of a red card in soccer. J Am Statist A 89: 1124–1127.

12 Wang S J and Ke H J (2004). Curbing gambling activities on the internet. IEEE Aero Elec Sys Mag 19: 33–35.

$E_0(t_1)$	μ 1	2	2.5	3	5
continuous (6)	56.89	38.91	33.04	28.51	17.88
discrete (8) $I_1 = I_2 = 0$	57.21	39.34	33.51	28.98	18.38
discrete (8) $I_1 = I_2 = 2$	57.76	40.03	34.21	29.69	18.98
discrete (8) $I_1 = I_2 = 4$	58.27	40.68	34.87	30.34	19.56
discrete (8) $I_1 = I_2 = 6$	58.74	41.29	35.50	30.97	20.13

Table 1: Expected time (minutes) of first goal, calculated at the start of the match using various different assumptions, Poisson means μ and expected added time periods I_1 and I_2

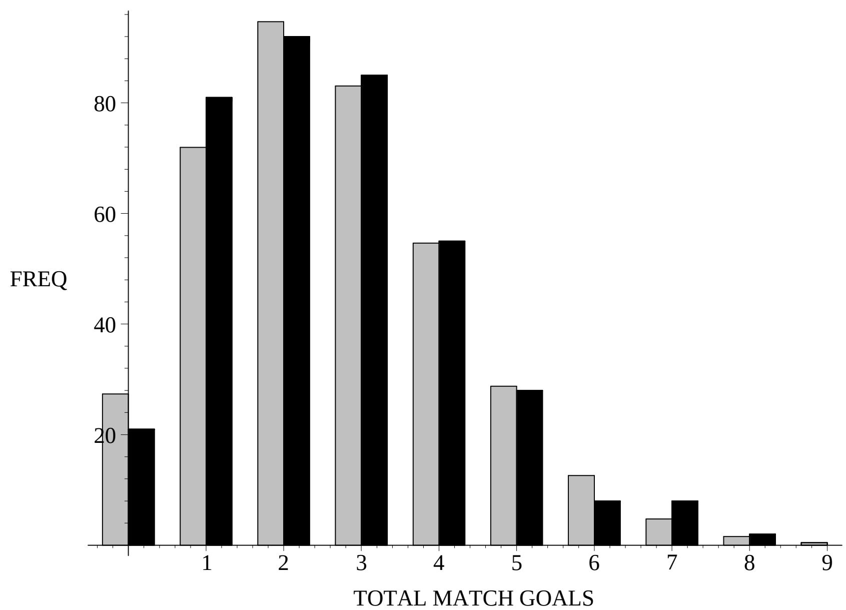

Figure 1: Fit to Poisson distribution for total match goals in all 380 Premiership matches for season 2002-3 ($\mu = 2.632$, actual frequency of scored goals in black, Poisson predictions shaded grey)

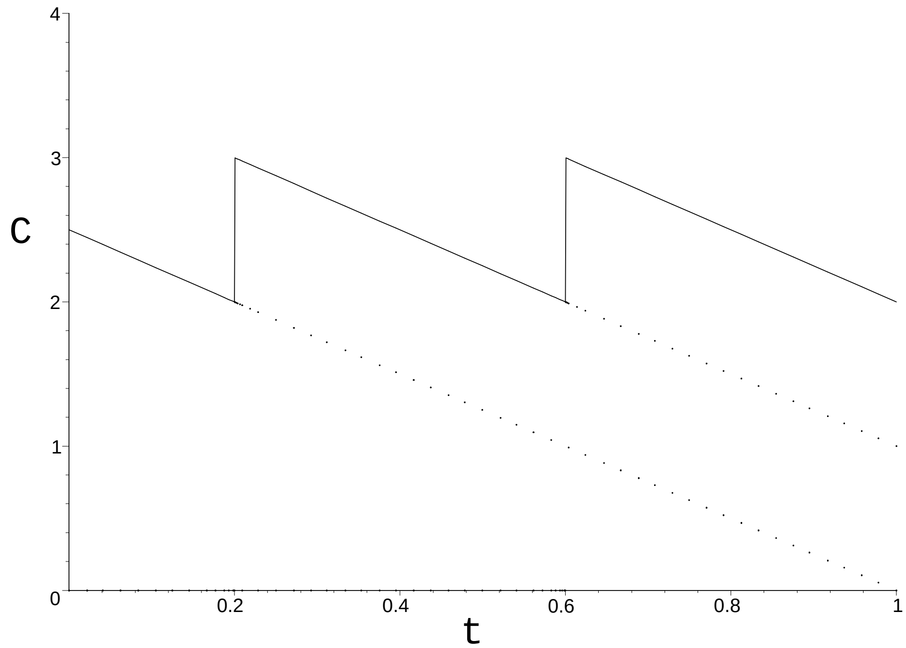

Figure 2: Centrespread C for the total number of goals during a match with $\mu = 2.5$ and goals at times 18mins and 54mins, i.e. $t_1 = \frac{18}{90} = 0.2$ and $t_2 = \frac{54}{90} = 0.6$. (Broken lines denote the valuation of the centrespread in the absence of any further events occurring)

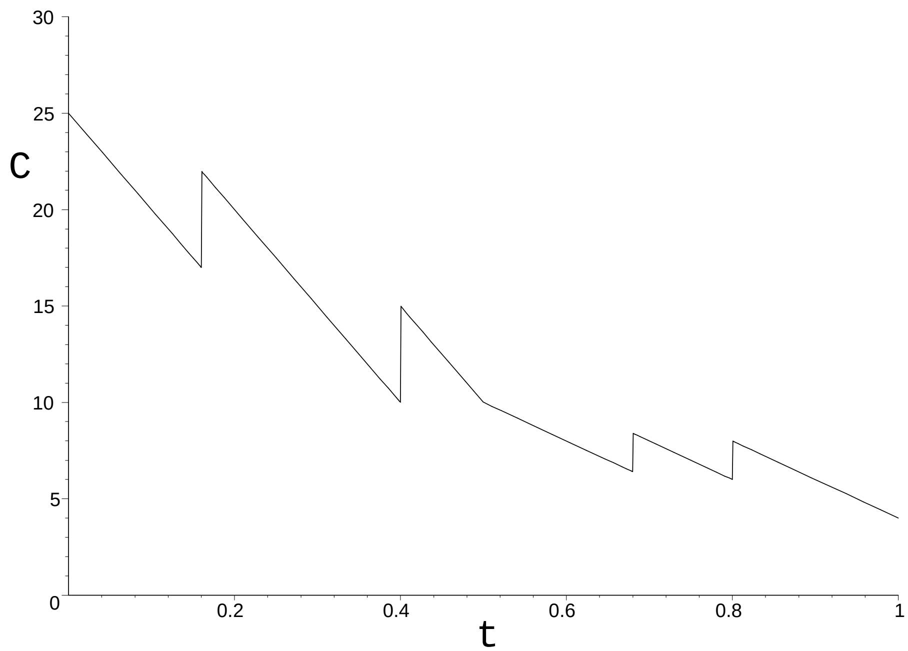

Figure 3: Centrespread C for multicorners market during a match with $\mu = 10$ and four corners at times $t_1 = 0.16,\ t_2 = 0.4,\ t_3 = 0.68$ and $t_4 = 0.8$

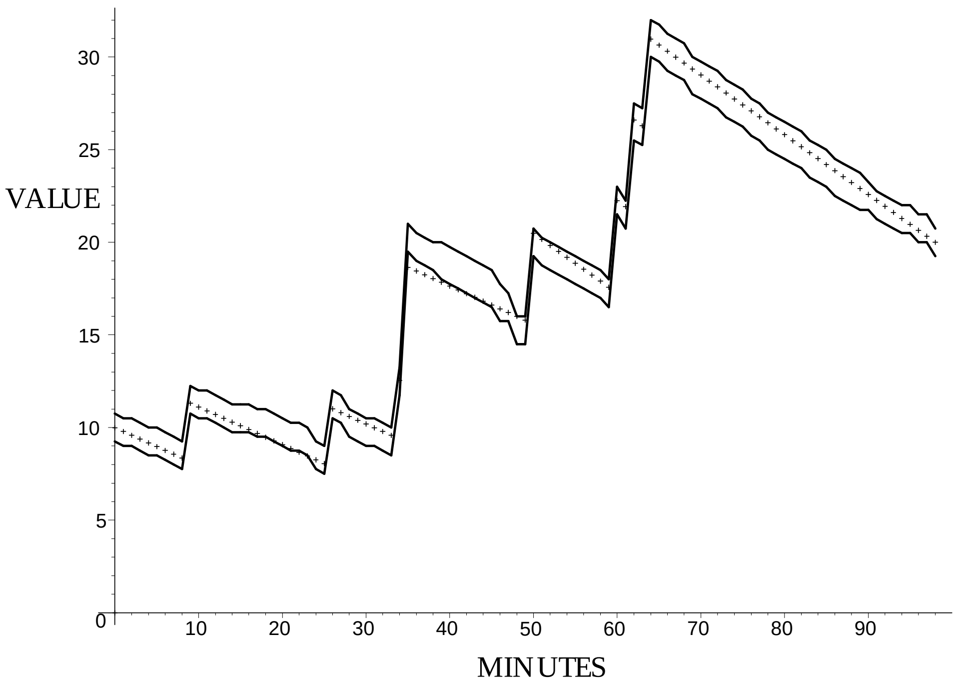

Figure 4: Comparison of centrespread prediction (3) (symbols) with OB's quoted dynamic spread (solid lines) for Portugal multicorners during the Portugal vs. Holland game in Euro2004. Note: the required Poisson parameters were inferred from the initial quoted spreads for other markets rather than fitted after the match

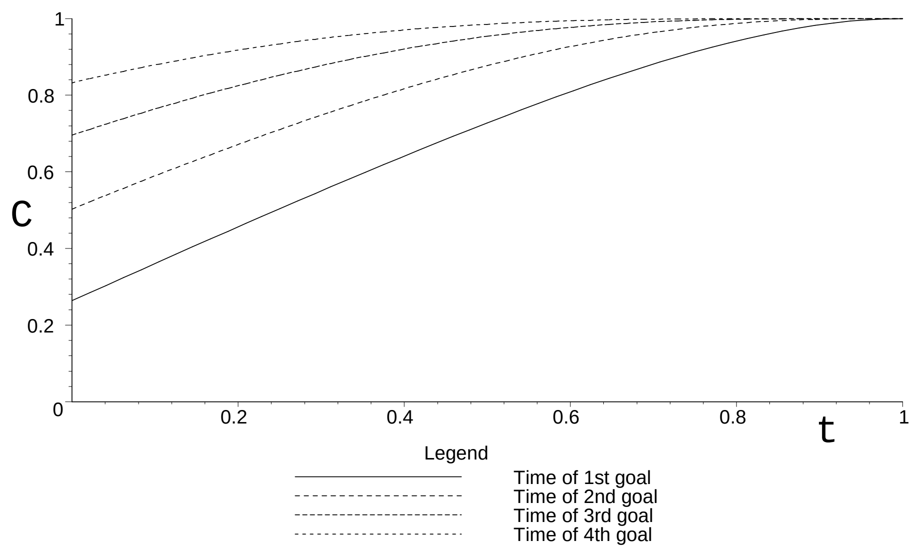

Figure 5: Centrespread C for the time of nth goal markets during a match with $\mu = 3.7$ and no goals scored

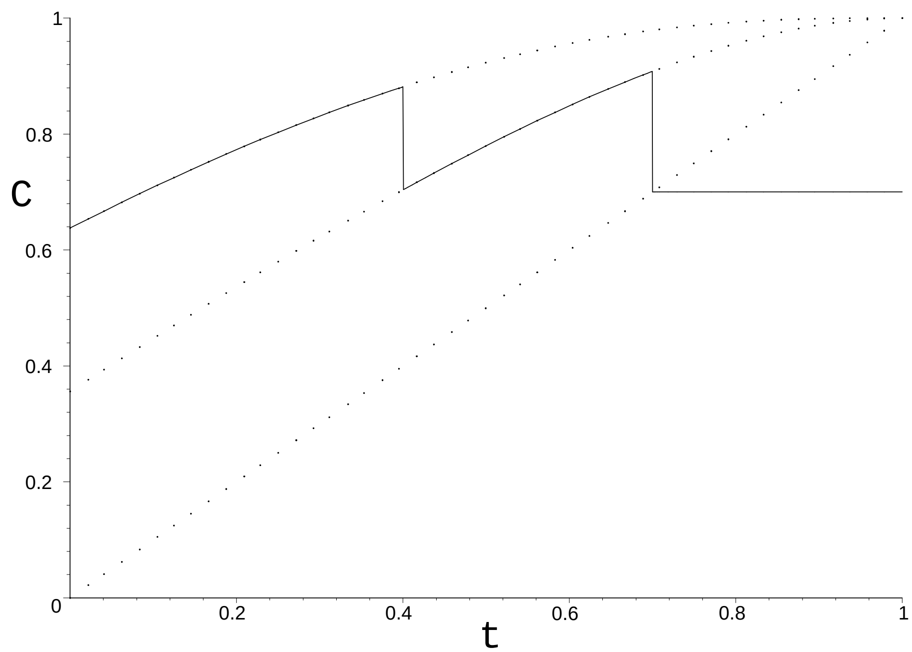

Figure 6: Centrespread for the time of 2nd goal market during a match with $\mu = 2.6$ and goals at times $t_1 = 0.4$ and $t_2 = 0.7$ (broken lines show how market evolves had further events not occurred - market ends at t_2 when 2nd goal is scored)

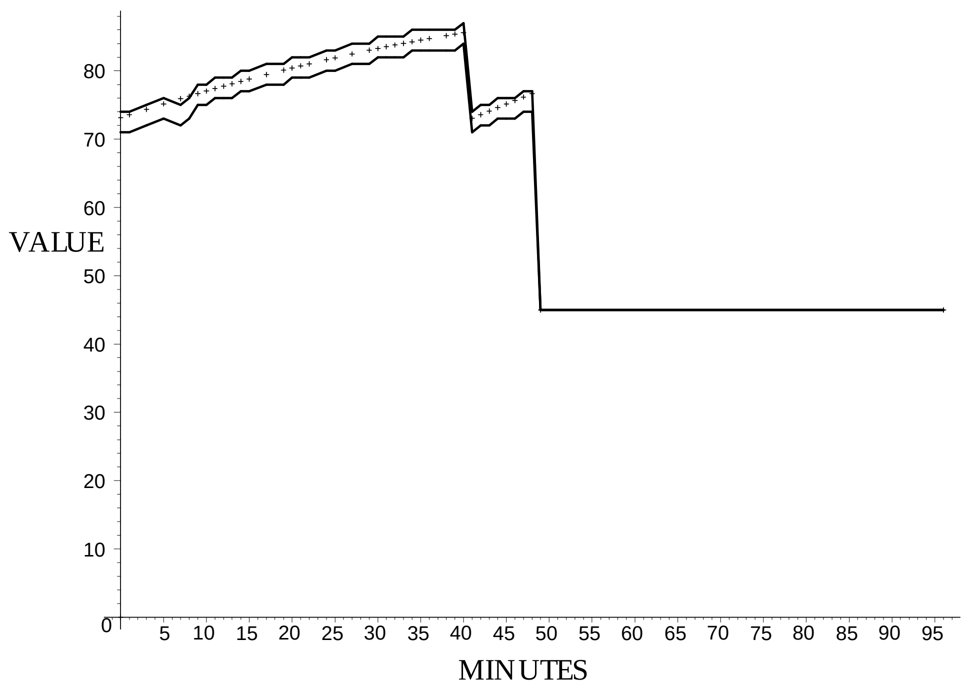

Figure 7: Comparison of centrespread prediction (6) (symbols) with OB's quoted dynamic spread (solid lines) for the time of England's 2nd goal during the England vs. Croatia game in Euro2004. Note: the required Poisson parameters were inferred from the initial quoted spreads for other markets rather than fitted after the match

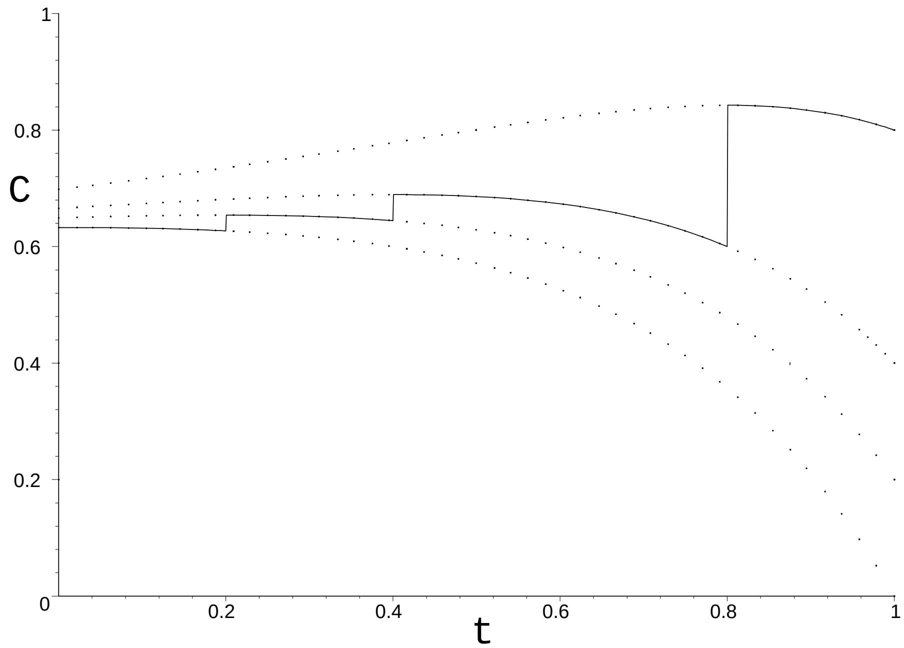

Figure 8: Centrespread for the time of last goal market during a match with $\mu = 2.5$ and goals at times $t_1 = 0.2$, $t_2 = 0.4$ and $t_3 = 0.8$. (Broken lines denote the valuation of the centrespread in the absence of any further events occurring)

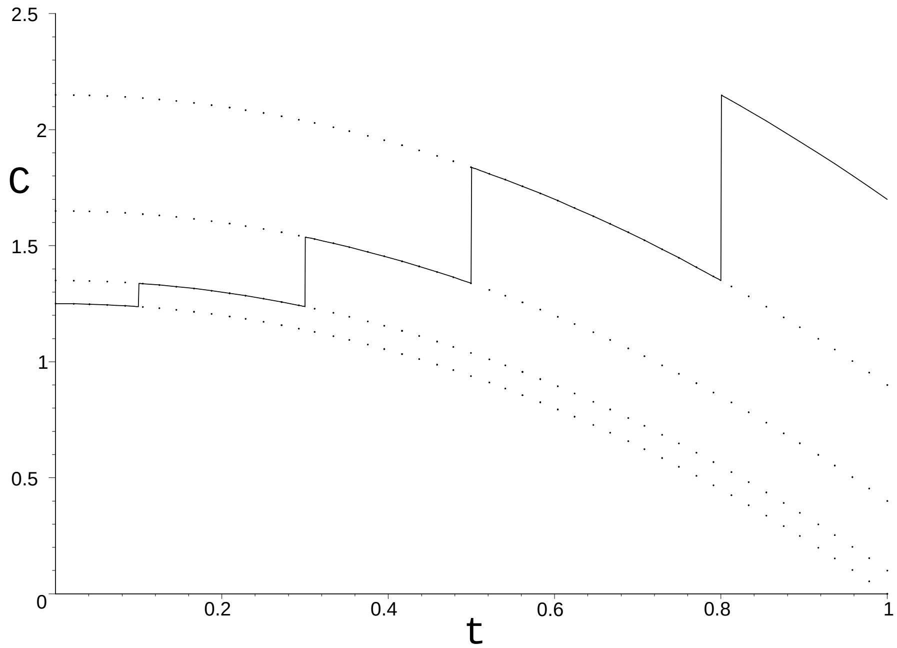

Figure 9: Centrespread for the total goal minutes market during a match with $\mu = 2.5$ and goals at times $t_1 = 0.1$, $t_2 = 0.3$, $t_3 = 0.5$ and $t_4 = 0.8$. (Broken lines denote the valuation of the centrespread in the absence of any further events occurring)

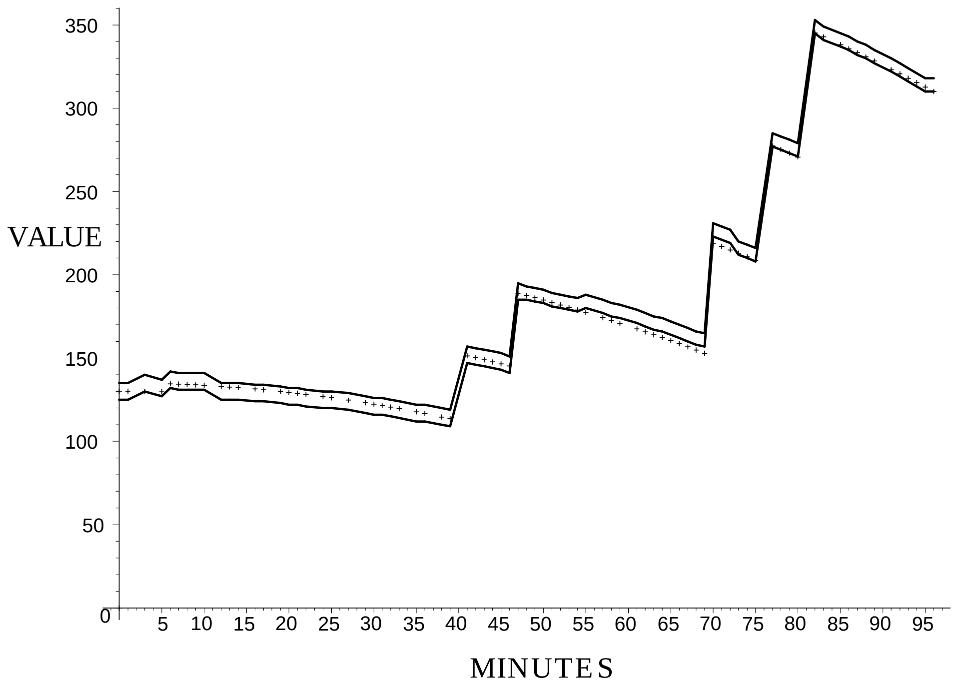

Figure 10: Comparison of centrespread prediction (6) (symbols) with OB's quoted dynamic spread (solid lines) for the total goal minutes market during the England vs. Croatia game in Euro2004. Note: the required Poisson parameters were inferred from the initial quoted spreads for other markets rather than fitted after the match

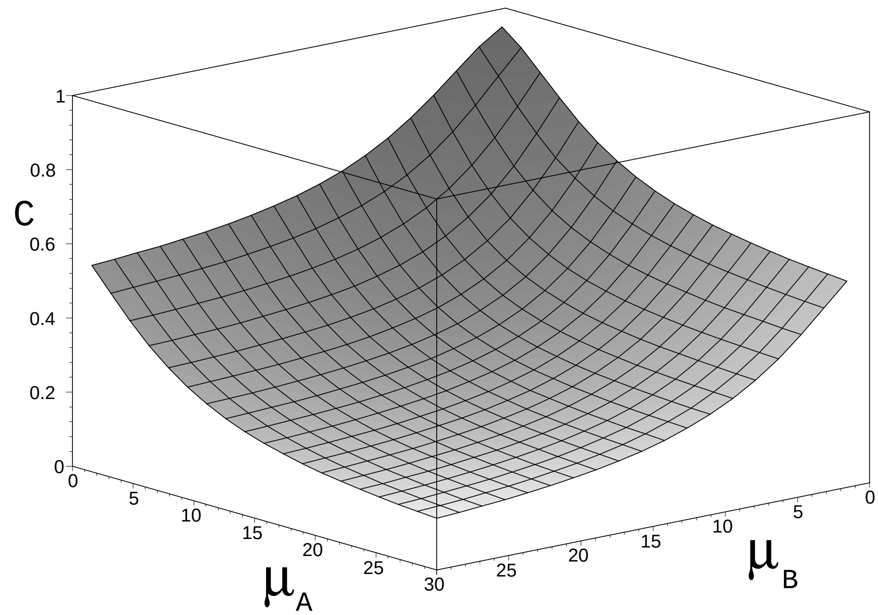

Figure 11: Centrespread at $t = 0$ for the four flags market as function of μ_A and μ_B

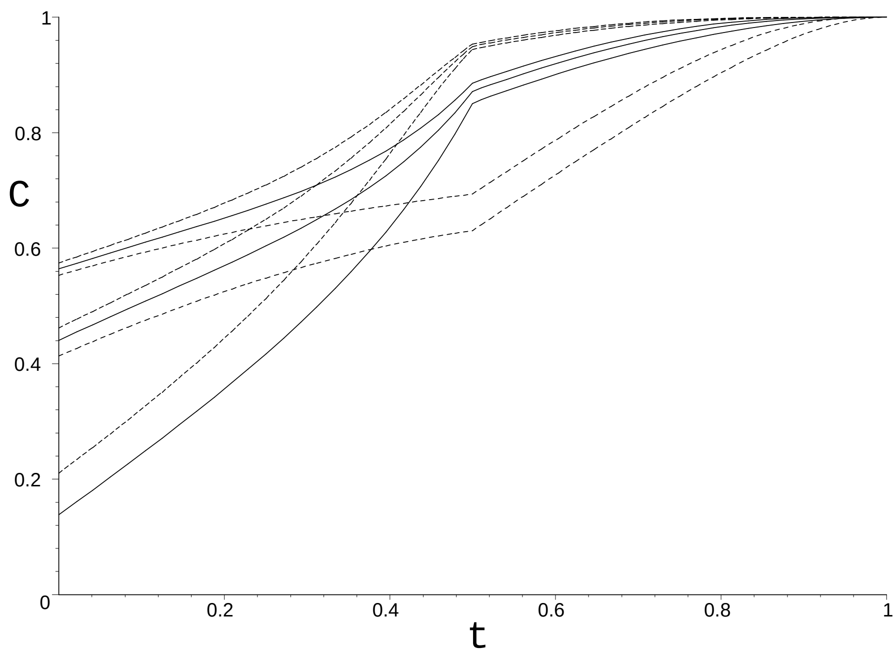

Figure 12: Possible centrespreads ("railtracks") for the four flags market with $\mu_A = 3$ and $\mu_B = 15$. Eight possible curves are shown, depending on the flags already used. At $t = 0$, the top curve is for $r_A = r_B = 2$, meaning that no corner has yet been taken. The second curve is for $r_A = 2,\ r_B = 1$, showing the C as a function of time when at least one corner has been taken from B_1 or B_2. The third curve is for $r_A = 2,\ r_B = 0$, (both B flags already used). The fourth curve represents the case $r_A = 1,\ r_B = 2$. The remaining curves in order cover the cases $r_A = r_B = 1$; $r_A = 1,\ r_B = 0$; $r_A = 0,\ r_B = 2$ and $r_A = 0,\ r_B = 1$, respectively

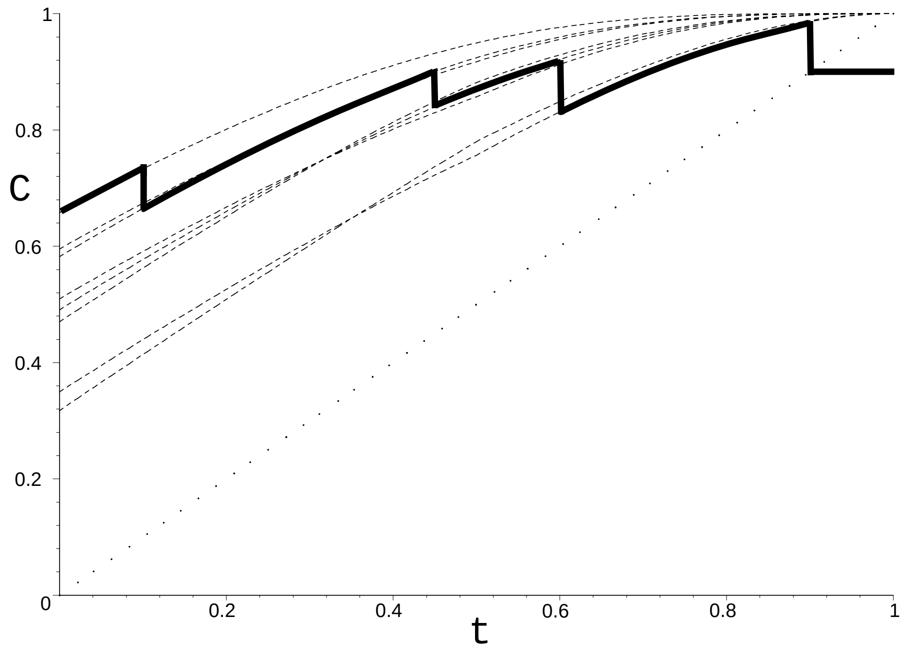

Figure 13: Evolving centrespread (bold line) for the four flags market with $\mu_A = 5.2$ and $\mu_B = 6.1$. The times of the first corners are $t_{A_1} = 0.1$, $t_{A_2} = 0.9$, $t_{B_1} = 0.45$ and $t_{B_2} = 0.6$. Since $\mu_A \sim \mu_B$, the shape is similar to the centrespread of the Time of nth goal market

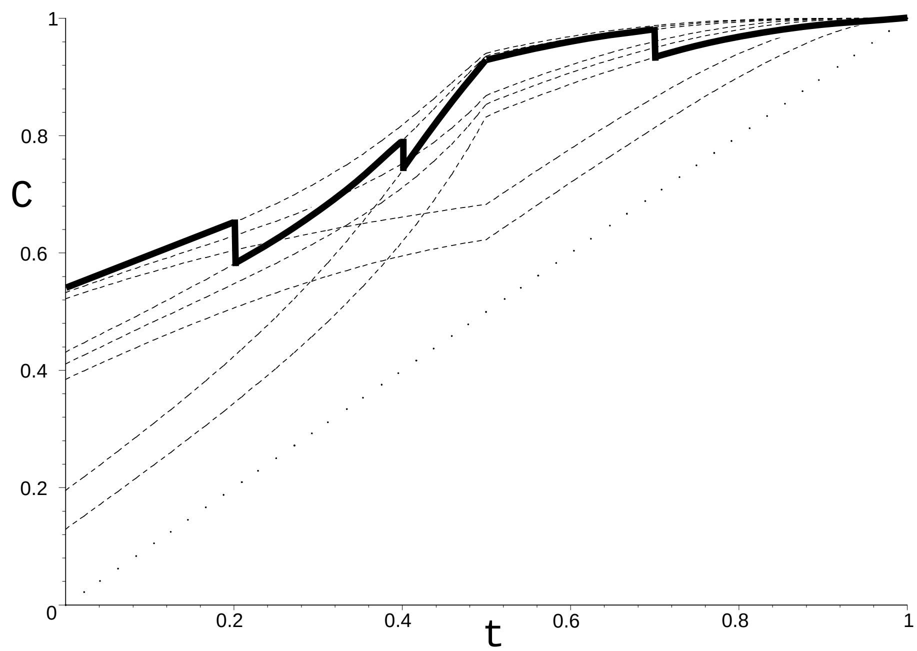

Figure 14: Evolving centrespread (bold line) for the four flags market with $\mu_A = 3.5$ and $\mu_B = 16$. The times of first corners are $t_{A_1} = 0.2$, $t_{A_2} = 0.4$, $t_{B_1} = 0.7$ and $t_{B_2} = 1$. The large difference between μ_A and μ_B leads to an exotic centrespread