The Market for EPL Odds

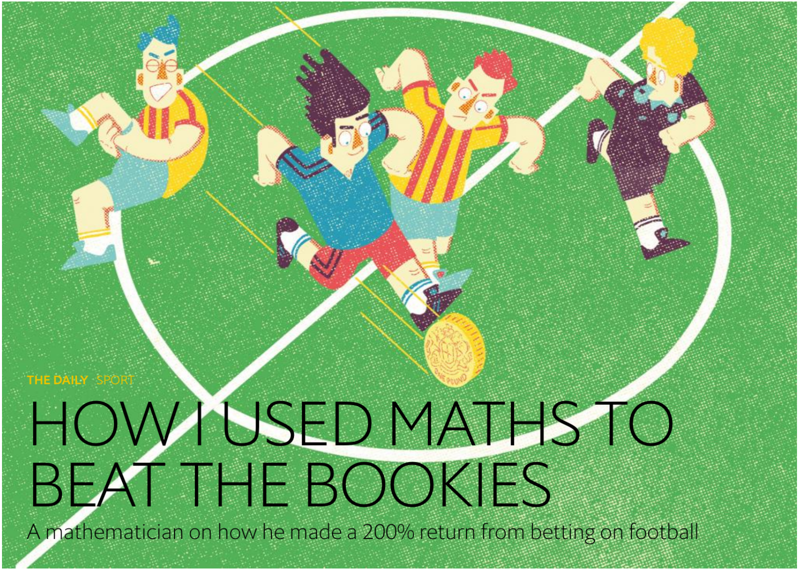
Guanhao Feng

Booth School of Business, University of Chicago

R/Finance 2017

(Joint work with Nicholas Polson and Jianeng Xu)

Soccermatics from David Sumpter



Underdog Leicester defeated odds of 5000-1 to win the Premier League

The Soccer Betting Market

- In 2013, gambling on soccer is a global industry worth anywhere between $700 billion and $1 trillion a year.
- Now, odds (fixed-odd beting) are set via online and updated in real time during the game (Betfair, Bet365, Ladbrokes).
- Bets can be placed on any outcome from the game.
- Some investment firms have developed their sports trading team.

Everton vs West Ham (March 5th, 2016)

Original odds data from Ladbrokes before the game:

Home \Away	0	1	2	3	4	5
0	11/1	12/1	28/1	66/1	200/1	450/1
1	13/2	6/1	14/1	40/1	100/1	350/1
2	7/1	7/1	14/1	40/1	125/1	225/1
3	11/1	11/1	20/1	50/1	125/1	275/1
4	22/1	22/1	40/1	100/1	250/1	500/1
5	50/0	50/1	90/1	150/1	400/1	
6	100/1	100/1	200/1	250/1		
7	250/1	275/1	375/1			
8	325/1	475/1				

What we can learn from the bookies?

- We provide a method for calibrating real-time market odds for the evolution of score difference for a soccer game.
- We rely on the odds market efficiency to calibrate a probability model and provide the market forecast of the final result.
- We provide an interpretation of the betting market and how it reveals the market expectation changes during the game.
- For future research, can we really beat the market (bookies)?

What we do?

- The Black-Scholes formula is a continuous-time model that describes the price of the option over time.
- We use a discrete-time probabilistic model to describe the evolution of score differences: Skellam process.
- Ex-ante, we show how to calibrate expected goal scoring rates using market-based odds information during the game.
- As the game evolves, we use the updated market odds to re-estimate the model.

Notation I

- We decompose the scores of each team as

$$\begin{cases} N_A(t) = & W_A(t) + W(t) \\ N_B(t) = & W_B(t) + W(t) \end{cases}$$

where $W_A(t)$, $W_B(t)$ and $W(t)$ are independent processes with $W_A(t) \sim Poisson(\lambda^A t)$, $W_B(t) \sim Poisson(\lambda^B t)$.

- Here $W(t)$ is a non-negative integer-valued process to induce a correlation between the numbers of goals scored.
- By modeling the score difference, we avoid having to specify the distribution of $W(t)$ as the score difference is independent of $W(t)$.

Notation II

- We define a Skellam process

$$N(t) = N_A(t) - N_B(t) = W_A(t) - W_B(t) \sim Skellam(\lambda^A t, \lambda^B t),$$

where $\lambda^A t$ is the cumulative expected scoring rate on the interval $[0, t]$.

- At time t, the conditional distributions for scores are

$$\begin{cases} W_A(1) - W_A(t) \sim & Poisson(\lambda^A(1-t)) \\ W_B(1) - W_B(t) \sim & Poisson(\lambda^B(1-t)) \end{cases}$$

- The conditional distribution for score difference is

$$N(1) - N(t) \sim Skellam(\lambda^A(1-t), \lambda^B(1-t)).$$

Conditional Probability Calculation

- Specifically, conditioning on $N(t) = \ell$, we have the identity

$$N(1) = \ell + Skellam(\lambda_t^A, \lambda_t^B)$$

- The probability of home team A winning at time t is

$$\begin{aligned} & \mathbb{P}(N(1) > 0 | \lambda_t^A, \lambda_t^B, N(t) = \ell) \\ = \; & \mathbb{P}(Skellam(\lambda_t^A, \lambda_t^B) > -\ell | \lambda_t^A, \lambda_t^B) \end{aligned}$$

- The probability of a draw at time t is

$$\begin{aligned} & \mathbb{P}(N(1) = 0 | \lambda_t^A, \lambda_t^B, N(t) = \ell) \\ = \; & \mathbb{P}(Skellam(\lambda_t^A, \lambda_t^B) = -\ell | \lambda_t^A, \lambda_t^B) \end{aligned}$$

Dynamic Model Calibration

Use real-time market odds to calibrate parameters λ_t^A and λ_t^B.

- Convert odds ratios to the implied probabilities of final scores

$$\mathbb{P}(N_A(1) = i, N_B(1) = j) = \frac{1}{1 + odds(i,j)}.$$

- The unconditional moments are given by

$$E[N(1)] = E[W_A(1)] - E[W_B(1)] = \lambda^A - \lambda^B,$$

$$V[N(1)] = V[W_A(1)] + V[W_B(1)] = \lambda^A + \lambda^B.$$

- The conditional moments are given by

$$\begin{cases} E[N(1)|N(t) = \ell] = \ell + (\lambda_t^A - \lambda_t^B), \\ V[N(1)|N(t) = \ell] = \lambda_t^A + \lambda_t^B. \end{cases}$$

Time-Varying Discussion

- Our approach: re-estimate $\{\lambda_t^A, \lambda_t^B\}$ dynamically through the real-time updated market odds.
- An alternative approach to time-varying $\{\lambda_t^A, \lambda_t^B\}$ is to use a Skellam regression with conditioning information such as possession percentages, shots (on goal), corner kicks, yellow cards, red cards, etc.
- We would expect jumps in the $\{\lambda_t^A, \lambda_t^B\}$ during the game when some important events happen. A typical structure takes the form

$$\begin{cases} \log(\lambda_t^A) = & \alpha_A + \beta_A X_{A,t-1} \\ \log(\lambda_t^B) = & \alpha_B + \beta_B X_{B,t-1}, \end{cases} \tag{1}$$

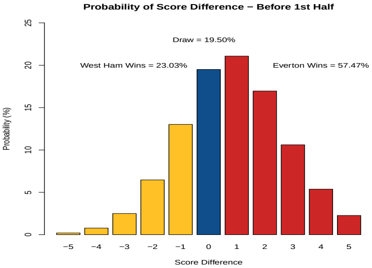
estimated using standard log-linear regression.

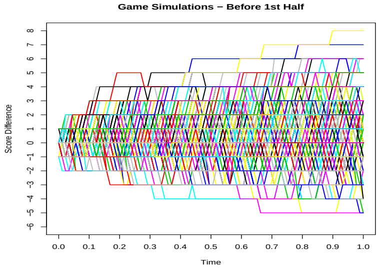
Everton vs West Ham (March 5th, 2016)

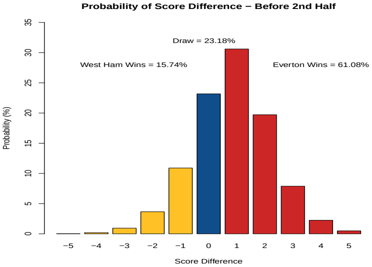
We use the market information for score difference.

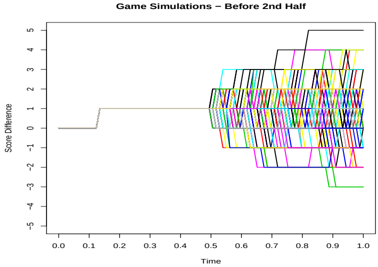
Home \Away	0	1	2	3	4	5
0	11/1	12/1	28/1	66/1	200/1	450/1
1	13/2	6/1	14/1	40/1	100/1	350/1
2	7/1	7/1	14/1	40/1	125/1	225/1
3	11/1	11/1	20/1	50/1	125/1	275/1
4	22/1	22/1	40/1	100/1	250/1	500/1
5	50/0	50/1	90/1	150/1	400/1	
6	100/1	100/1	200/1	250/1		
7	250/1	275/1	375/1			
8	325/1	475/1				

Market Implied Outcomes







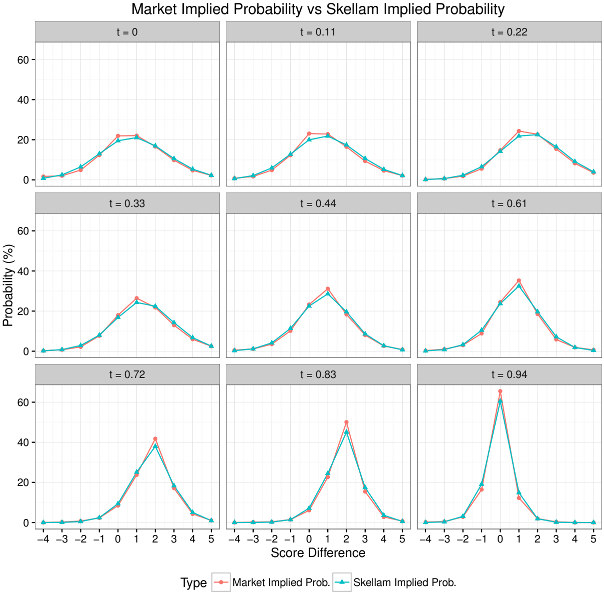


Skellam Process Approximation I

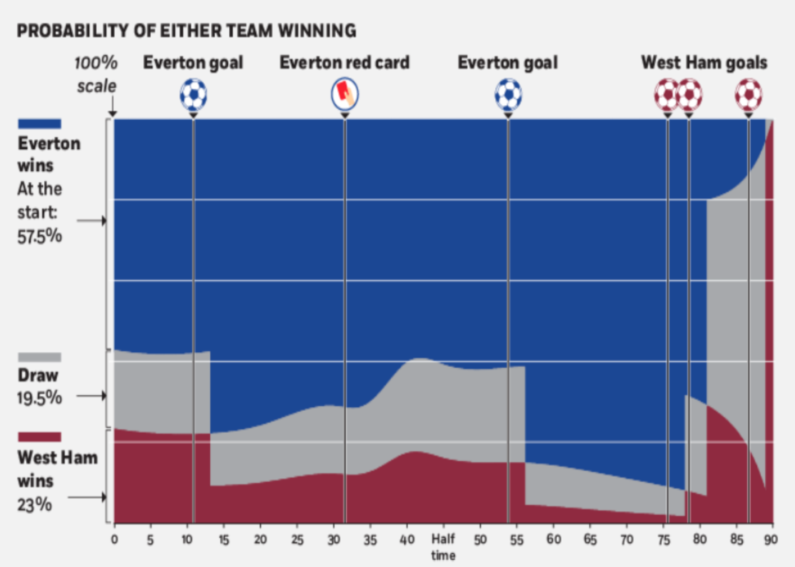
Market implied and Skellam implied probabilities for score differences before the game: ($\lambda_0^A = 2.33$, $\lambda_0^B = 1.44$)

Score difference	-4	-3	-2	-1	0
Market Prob. (%)	1.70	2.03	4.88	12.33	21.93
Skellam Prob.(%)	0.78	2.50	6.47	13.02	19.50
Score difference	1	2	3	4	5
Market Prob. (%)	22.06	16.58	9.82	4.72	2.23
Skellam Prob.(%)	21.08	16.96	10.61	5.37	2.27

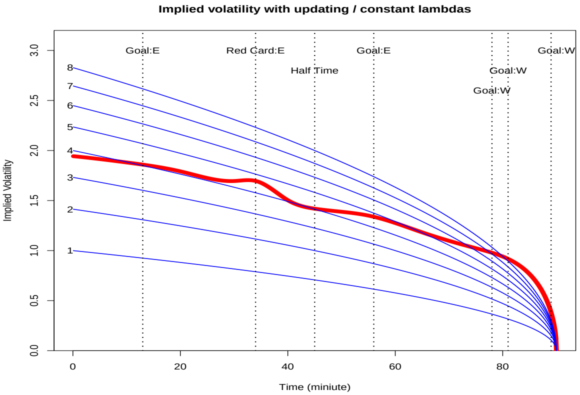
Skellam Process Approximation II



Win/Draw/Lose Probability Evolution



Odds Implied Volatility



- We define a discrete version of the implied volatility of the games outcome as $\sigma_{IV,t} = \sqrt{\lambda_t^A + \lambda_t^B}$.
- Red line: the path of implied volatility: $\sigma_t^{red} = \sqrt{\hat{\lambda}^A(1-t) + \hat{\lambda}^B(1-t)}$.
- Blue reference lines: constant volatility $\lambda^A + \lambda^B$: $\sigma_t^{blue} = \sqrt{(\lambda^A + \lambda^B) * (1-t)}$.