Joint modelling of goals and bookings in association football matches

Andrew Titman
Debbie Costain   Gareth Ridall   Kris Gregory

Lancaster University

September 4, 2012

Overview

- Data
- Counting process model for goals and bookings
- Modelling effects of covariates and game events
- Results
- Validation via live spread betting

Data

- BBC live text reports on 1,864 Premier League and Championship games from the 2009/10 and 2010/11 seasons
  - Data on time of goals, yellow cards and red cards to nearest second in relation to time since kickoff (or time since half-time)
  - e.g. away yellow card at 32:36, home goal at 93:23.
- Additional covariate information:
  - Match referee
  - Bookmaker's prior match outcome odds
  - Bookmaker's prior odds of at least 3 goals in the game.

Aims

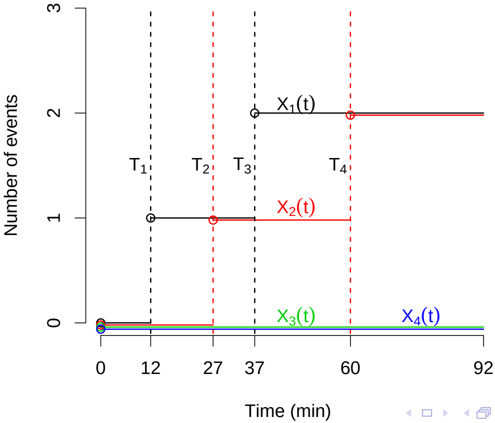
- Determine the factors affecting incidence of bookings within a match
- Determine the extent to which bookings affect goal scoring intensities
- Allow dynamic prediction of match outcomes
- Extending previous work on modelling football
  - e.g. Dixon and Robinson (1998): Poisson birth process models for goals.
  - e.g. Buraimo et al (2010): Factors affecting incidence of bookings conditional on current match status.

Models for football

- The processes of accrual of goals, yellow cards and red cards can be modelled as counting processes

$$
\begin{aligned}
X_1(t) &= \#\{\text{Home goals by time } t\} \\
X_2(t) &= \#\{\text{Away goals by time } t\} \\
X_3(t) &= \#\{\text{Home yellow cards by time } t\} \\
X_4(t) &= \#\{\text{Away yellow cards by time } t\} \\
X_5(t) &= \#\{\text{Home straight red cards by time } t\} \\
X_6(t) &= \#\{\text{Away straight red cards by time } t\} \\
X_7(t) &= \#\{\text{Home reds from 2nd offence by time } t\} \\
X_8(t) &= \#\{\text{Away reds from 2nd offence by time } t\ \}
\end{aligned}
$$

Models for football



Models for football

- Associated intensity for each counting process:

$$\lambda_j(t; \mathcal{F}_{t-}, z) = \lim_{\delta t \downarrow 0} \frac{\mathbb{E}\{X_j(t + \delta t) - X_j(t) | \mathcal{F}_{t-}, z\}}{\delta t}$$

  $\mathcal{F}_{t-}$ denotes filtration of history of whole multivariate process $\mathbf{X}(t)$ up to $t-$.

- Simplifying (Markov) assumption:

$$\lambda_j(t; \mathcal{F}_{t-}, z) = \lambda_j(t; \mathbf{X}(t-), z)$$

Modelling the effect of game events

- We make a proportional intensities assumption where

$$\lambda_j(t; \mathbf{X}(t-), z) = \lambda_j(t) \exp\{f(\mathbf{X}(t-), z; \beta)\}$$

  for some function f with associated regression parameters β including time fixed covariates z
- For instance, we allow goal scoring rates to depend on the current score line, creating separate factor levels e.g. $I(X_1(t) = X_2(t))$, $I(X_1(t) = 1 \cap X_2(t) = 0)$, $I(X_1(t) > 1 \cap X_1(t) > X_2(t))$ etc.
- Baseline intensities taken to be of Weibull form
  - All event intensities increase as match progresses

Accounting for team ability

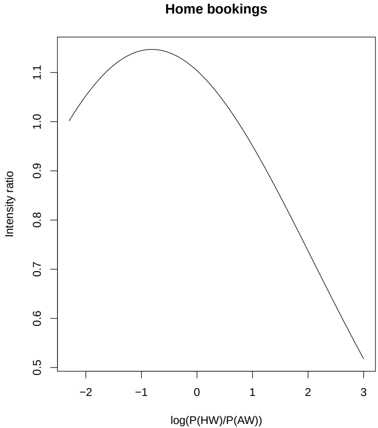
- Goal scoring rates will clearly depend on relative team abilities
- Use bookmaker's prior odds of match outcomes
  - To inform both goal scoring rates and booking rates
  - Included in model as B-spline functions of the implied $\log\{P(\text{Home Win})/P(\text{Away Win})\}$
  - Bookmaker's odds of at least 3 goals in a game used to further adjust rates.

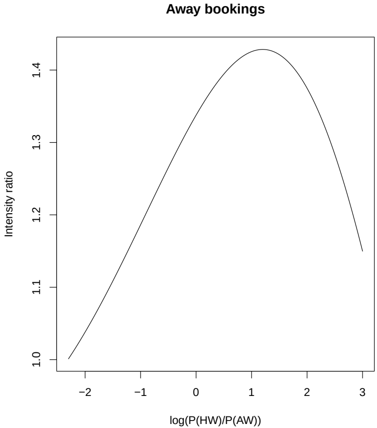
Results: Goal scoring rates

- Strong dependence of team ability on scoring rates.
- Scoring rate of both teams decreases once there is a score draw (by 13% for the home team and 27% for the away team compared to their rates when it is 0-0).
- Some evidence that teams 'sit back' once they are ahead
- No direct effect of yellow cards on goal scoring rates
- Red cards:
  - Home Red card $\rightarrow$ 60% increase in away team's scoring rate, 17% decrease in home team scoring rate.
  - Away Red card $\rightarrow$ 69% increase in home team's scoring rate, 42% decrease in away team scoring rate.
  - Handicap for away team appears more severe.

Results: Booking rates

- Significant dependence of team ability on booking rates





Results: Booking rates

- No substantial dependence of score line on bookings after accounting for team ability.
- Escalation effects:
  - A team's booking rate increases by 25% if the opposing team get a yellow card.
  - A yellow card to any player on a team more than doubles the hazard of a straight red card to any other player on that team. (ie. referees rarely give a red card as the first bookable offence in the game)
  - Hazard of red cards from a second offence is close to proportional to the number of yellow cards already awarded to that team.
- Lower rate of bookings in the Championship than Premier League (11% lower)

Spread betting

- Spread betting allows one to bet on a range of match outcomes
  - e.g. total number of goals
  - e.g. total booking points (yellow card = 10, red card = 25, yellow + red card = 35)
  - e.g. goal difference
- Can either "buy" :a bet that the outcome will be higher than the available "price"
- or "sell' :a bet that the outcome will be lower than the available "price"

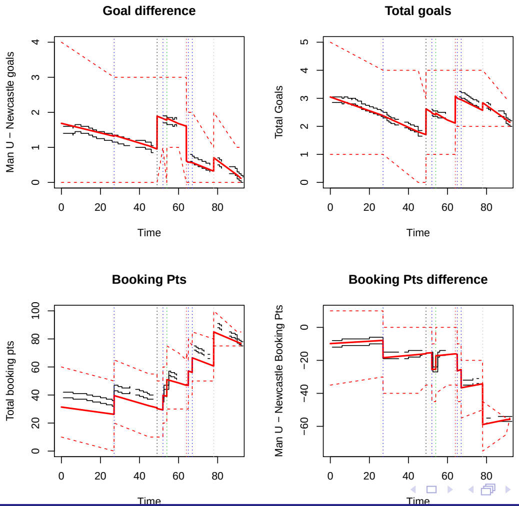
Betting strategy

- Broadly want to maximize expected pay out
- but allow for some degree of risk aversion: i.e. how much more do we prefer a £1 gain with no risk to a £1 expected gain with variance of 1.
- Here consider a utility function for the gain:

$$U(G(\mathbf{X};\mathcal{A})) = 1 - \exp\{-\kappa G(\mathbf{X};\mathcal{A})\}$$

  where $\kappa > 0$ corresponds to risk aversion and $G(\mathbf{X};\mathcal{A})$ is the gain associated with match outcome $\mathbf{X}$ and betting action $\mathcal{A}$.
- Choose the action $\mathcal{A}$ (e.g. buy or sell x) that maximizes the expected utility
- Model predictions computed through simulation at each minute of the match

Example: Man Utd v Newcastle, 26th November 2011



Overall performance

- 93 matches monitored (from *sportingindex.com*) and betting strategy applied retrospectively (i.e. **no actual money wagered!**)

Market	Overall gain
Goal Difference	-2.3
Total Goals	-18.2
Total Bookings	38.9
Booking Pt Difference	14.8
Total	32.7

- Most gain is from bookings markets
- Poorer performance for total goals
- Larger variation in outcomes for individual games

Further Issues

- Referee effects: There is inconsistency across individual referees in the number of cards issued and evidence this is already accounted for in bookmaker's prices.
- Team characteristics, independent of abilitiy, also important (aggressiveness measured by number of fouls in previous matches).
- High profile matches have higher booking rates (e.g. matches between top 5 Premier League clubs)
- Further work
  - Better model for goal scoring rates, e.g. take into account individual club's past scoring record.
  - More realistic utility functions, e.g. log-utility, taking into account total wealth at particular time.

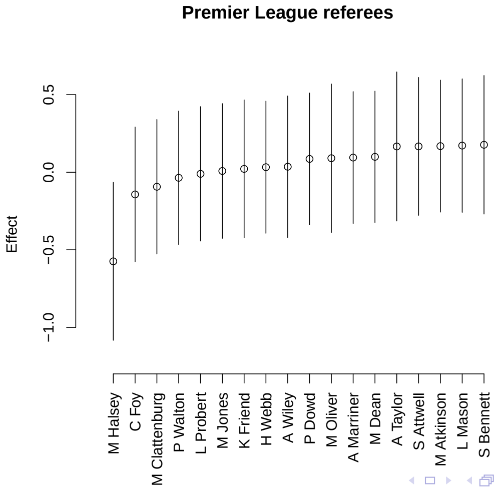
References

- Dixon MJ, Robinson ME. A birth process model for association football matches. The Statistician 1998; 47: 523-538.
- Buraimo B, Forrest D, Simmons R. The 12th man?: referring bias in English and German soccer. JRSS A 2010; 173: 431-449.

Non-Markov effects

- Red cards (and to a lesser extent yellow cards) are often associated with a free-kick or penalty implying a goal scoring opportunity for the opposing team
- Subsequent effect of the red card is then a mixture of short term effects related to the penalty and persistent effects.
- Ideally would have data on whether the booking resulted in a direct goal scoring opportunity and would model this as a separate event.
- To investigate this can create a time dependent covariate:
  - Takes value 1 for three minutes after a red card, 0 otherwise.
- Then look at the interaction of the this covariate and # Red cards.

Referee Effects: Premier League



Referee Effects: Championship