

Profiting from an inefficient association football gambling market: Prediction, risk and uncertainty using Bayesian networks



Anthony Costa Constantinou*, Norman Elliott Fenton, Martin Neil

Risk & Information Management (RIM) Research Group, Department of Electronic Engineering and Computer Science, Queen Mary, University of London, London E1 4NS, United Kingdom

ARTICLE INFO

Article history:

Received 4 December 2012
Received in revised form 20 May 2013
Accepted 21 May 2013
Available online 4 June 2013

Keywords:

Bayesian networks
Expert systems
Football betting
Football forecasts
Subjective information

ABSTRACT

We present a Bayesian network (BN) model for forecasting Association Football match outcomes. Both objective and subjective information are considered for prediction, and we demonstrate how probabilities transform at each level of model component, whereby predictive distributions follow hierarchical levels of Bayesian inference. The model was used to generate forecasts for each match of the 2011/2012 English Premier League (EPL) season, and forecasts were published online prior to the start of each match. Profitability, risk and uncertainty are evaluated by considering various unit-based betting procedures against published market odds. Compared to a previously published successful BN model, the model presented in this paper is less complex and is able to generate even more profitable returns.

© 2013 The Authors. Published by Elsevier B.V. Open access under [CC BY license](#).

1. Introduction

Association Football (hereafter referred to as simply *football*) is the most popular sport internationally [10,27,11], and attracts an increasing share of the multi-billion dollar gambling industry; particularly after its introduction online [6]. This is one of the primary reasons why we currently observe extensive attention paid to football odds by both academic research groups and industrial organisations who look to profit from potential market inefficiencies. While numerous academic papers exist which focus on football match forecasts, only a few of them appear to consider profitability as an assessment tool for determining a model's forecasting capability.

Pope and Peel [30] evaluated a simulation of bets against published market odds in accordance with the recommendations of a panel of newspapers experts. They showed that even though there was no evidence of abnormal returns, there was some indication that the expert opinions were more valuable towards the end of the football season. Dixon and Coles [8] were the first to evaluate the strength of football teams for the purpose of generating profit against published market odds with the use of a time-dependent

Poisson regression model based on Maher's [26] model. They formed a simple betting strategy for which the model was profitable at sufficiently high levels of discrepancy between the model and the bookmakers' probabilities. However, these high discrepancy levels returns were based on as low as 10 sample values; at lower discrepancy levels and with a larger sample size the model was unprofitable. The authors suggested that for a football forecast model to generate profit against bookmakers' odds without eliminating the in-built profit margin, "it requires a determination of probabilities that is sufficiently more accurate from those obtained by published odds". A similar paper by Dixon and Pope [9] was also published on the basis of 1993–1996 data and reported similar results. Rue and Salvesen [32] suggested a Bayesian dynamic generalised linear model to estimate the time-dependent skills of all the teams in the English Premier League (EPL) and English Division 1. They assessed the model against the odds provided by Intertops, a firm which is located in Antigua in the West Indies, and demonstrated profits of 39.6% after winning 15 bets out of a total of 48 for EPL matches, and 54% after winning 27 bets out of a total of 64 for Division 1 matches.

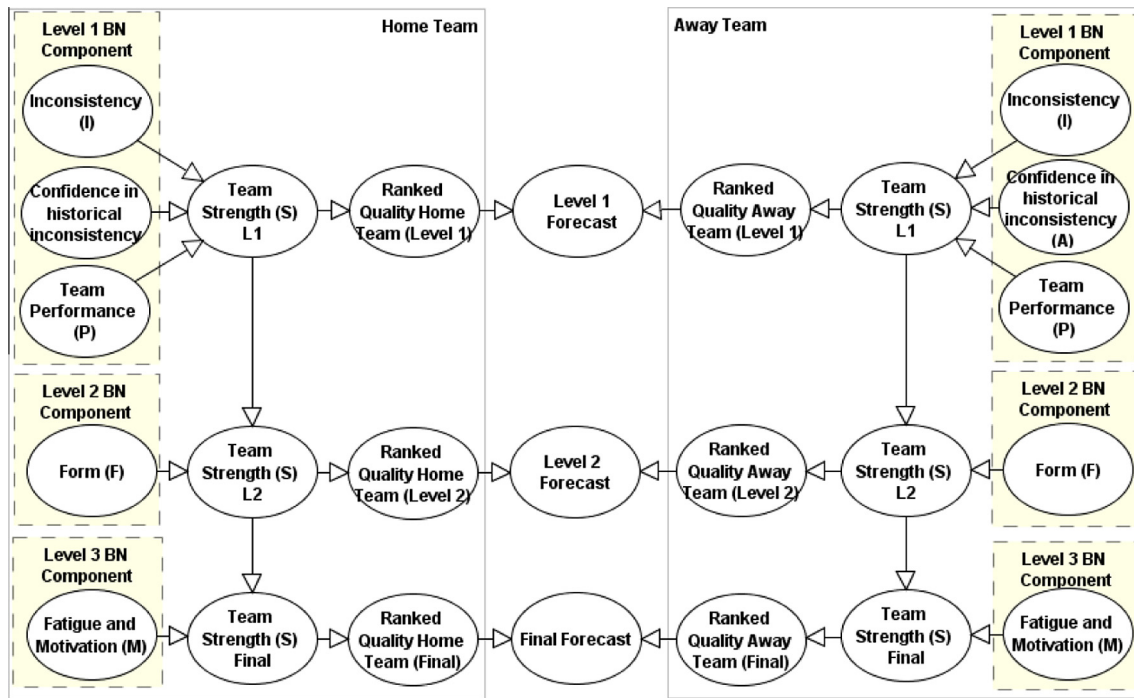
In an attempt to exploit the favourite-longshot bias for profitable opportunities, Poisson and Negative Binomial models have been used to estimate the number of goals scored by a team [3]. The conclusion was that even though the fixed odds offered against particular score outcomes did seem to offer profitable betting opportunities in some cases, these were few in number. Goddard and Asimakopoulos [17] proposed an ordered probit regression model to forecast EPL match results in an attempt to test the

* Corresponding author.

E-mail addresses: anthony@constantinou.info (A.C. Constantinou), norman@eeecs.qmul.ac.uk (N.E. Fenton), martin@eeecs.qmul.ac.uk (M. Neil).

Table 1How $S \rightarrow S_R$ is defined in 14 predetermined ranks, based on [7].

S	>89	85–89	80–84	75–79	70–74	...(intervals of 5 points)	25–29	20–24	<20
S_R	1	2	3	4	5	...	12	13	14

**Fig. 1.** Simplified model topology of the overall Bayesian network.

weak-form efficiency of prices in the fixed-odds betting market. To evaluate the model they considered seasons 1999 and 2000. Even though they reported a loss of -10.5% for overall performance, the model appeared to be profitable (on a pre-tax gross basis) at the start and at the end of every season.¹ Using a benchmark statistical model with a large number of quantifiable variables relevant to match outcomes Forrest et al. [15] examined the effectiveness of forecasts based on published odds and forecasts generated. They considered five different bookmaking firms for five consecutive seasons (1998–2003) and demonstrated that the model generated negative returns ranging from -10% to -12% depending on the bookmaking firm, but the loss was reduced to -6.6% when using the best available odds by exploiting arbitrage between bookmaking firms.

[19] attempted to investigate the rationality of bookmakers' odds using an ordered probit model to generate predictions for EPL matches. By considering William Hill odds, they followed the betting strategy introduced in [8,9] and reported negative returns ranging from -2.5% to -15% for all discrepancy levels during seasons 2004–2006. In the absence of any consistently successful model against market odds, the authors claimed that "if it was successful, it would not have been published". [21] considered the ELO rating system for football match prediction, although it was initially developed by [12] for assessing the strength of international chess players. Even though the ratings appeared to be useful in encoding the information of past results for measuring the strength of a team, resulting forecasts reported negative expected returns against numerous seasons of published odds using

various betting strategies. However, Constantinou and Fenton [5] later developed a novel rating technique (called pi-rating) that outperformed considerably the two ELO rating variants of [21], in terms of profitability, over a period of five EPL season.

[7] recently presented a Bayesian network model that was used to generate forecasts about the EPL matches during season 2010/2011, by considering both objective and subjective information for prediction. Forecasts were published online [29] prior to the start of each match, and this was the first academic study to demonstrate profitability that was consistent against published market odds over a sufficiently high number of betting trials without eliminating the bookmakers' profit margin.

In this paper we present a Bayesian network model for forecasting football outcomes that is based on the approach in [7], but with reduced complexity and higher forecasting capability (which we explain in detail in Sections 2–4). The paper is organised as follows: Section 2 describes the model; Section 3 presents the various betting procedures along with a Bayesian network component for assessing the risks involved under each of the procedures; Section 4 discusses the results; Section 5 provides our concluding remarks.

2. The model

In this section we first provide a brief overview of the model summarising the main differences to the approach in [7]. We then describe the technical components of the model in subsections.

As in [7] we have used the AgenaRisk Bayesian network tool to build the model. The most important differentiator between AgenaRisk and other Bayesian network tools is its ability to properly incorporate continuous variables, without any constraint, and without the need for static discretisation. It does this through

¹ Gross pre-taxed returns of $+3.1\%$ and $+1.5\%$ for respective seasons beginning 1999 and 2000, and gross returns of $+8\%$ for respective seasons ending 1999 and 2000.

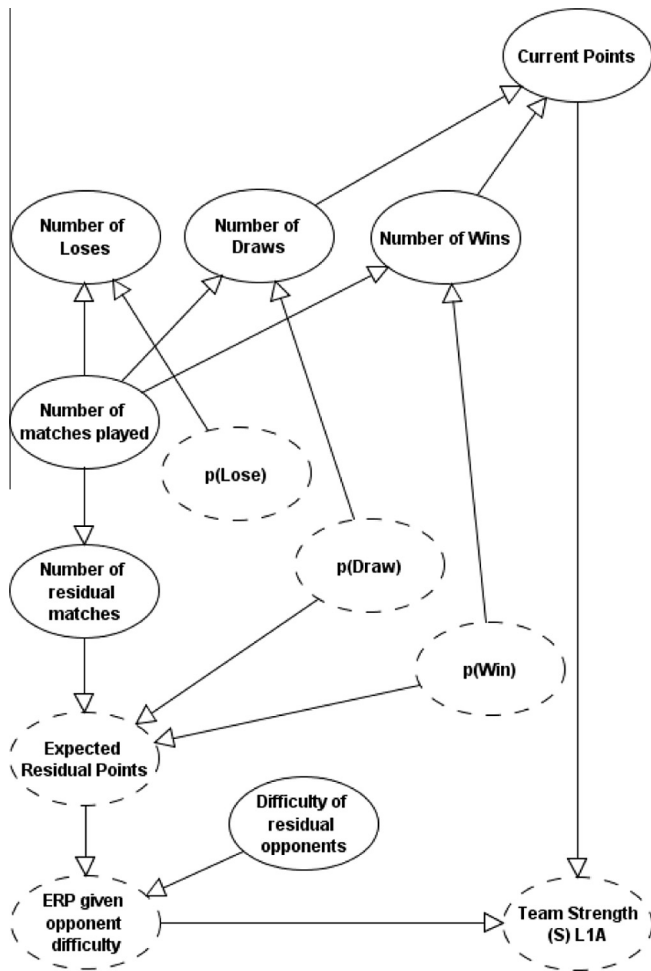


Fig. 2. Level 1 Component (P): formulating the S prior. Fig. G.2 presents an example of this component with different scenarios. Dashed nodes indicate latent variables. The variables *Current Points*, *Number of matches played* and *Number of residual matches* are definitional variables.

its revolutionary dynamic discretisation algorithm that produces results with far greater accuracy than is possible otherwise. The dynamic discretisation algorithm [28] uses entropy error as the basis for approximation. In outline, the algorithm follows these steps:

1. Convert the BN to a Junction Tree (JT) and choose an initial discretisation for all continuous variables.
2. Calculate the Node Probability Table (NPT) of each node given the current discretisation.
3. Enter evidence and perform global propagation on the JT, using standard JT algorithms.
4. Query the BN to get posterior marginals for each node, compute the approximate relative entropy error, and check if it satisfies the convergence criteria.
5. If not, create a new discretisation for the node by splitting those intervals with highest entropy error.
6. Repeat the process by recalculating the NPTs and propagating the BN, and then querying to get the marginals and then split intervals with highest entropy error.
7. Continue to iterate until the model converges to an acceptable level of accuracy.

This dynamic discretisation approach allows more accuracy in the regions that matter and incurs less storage space over static discretisations. In the implementation of the algorithm the user can select the number of iterations and convergence criteria, and

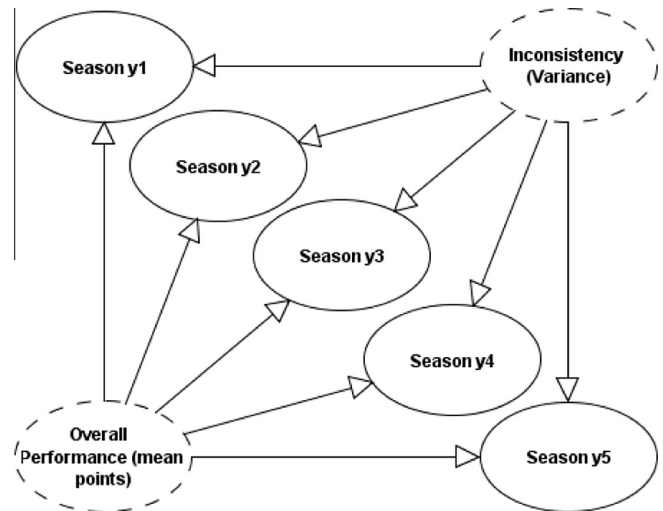


Fig. 3. Level 1 Component (I): measuring a team's *Inconsistency (Variance)* (V) based on league point totals over the five most recent seasons. Fig. G.3 presents an example of this component with different scenarios. Dashed nodes indicate latent variables.

hence can go for arbitrarily high precision (at the expense of increased computation times). Details about the role of qualitative judgments and how inference is done are provided in [13,28].

The model is constructed on the basis of three generic factors: *team strength*, *form*, and *fatigue with motivation*. There are model components corresponding to each of the three factors. The components are inferred hierarchically and at each level of hierarchy a match forecast is generated. This helps us understand how the probabilities transform at each level and determine the effectiveness of each model component by assessing the probability distributions generated at each level. Specifically:

1. At level 1, match forecasts of type $\{p(H), p(D), p(A)\}^2$ are generated based on each team's strength (S), where an S prior is formulated according to observed and expected results (P) of relevant match instances of the current season, and team inconsistencies (I) given relevant final league point totals from the five most recent seasons;
2. At level 2, posterior predictive distributions of S (from level 1) are formulated based on team form (F);
3. At level 3, posterior predictive distributions of S (from level 2) are formulated based on team fatigue and motivation (M).

Thus, the model follows hierarchical levels of Bayesian inference such that $S_1 \rightarrow S_2 \rightarrow S_3$, where $S_1 = p(S|P, I)$, $S_2 = p(S|S_1, F)$, and $S_3 = p(S|S_2, M)$.

The variable S is a $\sim TNormal(\mu, \sigma, 0, 114)^3$ probability density function and at each level of hierarchy represents a prediction of the team's strength which is measured in total league points. The distribution of S is summarised in 14 predetermined ranks (S_R) as presented in Table 1, whereby the granularity of the 14 ranks ensures that, for any match combination of parameters S_R , sufficient data points exist for a reasonably well informed match forecast prior. In particular, match forecasts given S_R are formulated on the basis of relevant historical match outcomes.⁴ Hence, the underlying

² Corresponding to home win, draw, and away win.

³ Truncated Normal where the endpoints are the respective minimum and maximum number of points a team can accumulate in an EPL season (38 games with three points for a win).

⁴ The database consists of the home, draw and away results of all the EPL matches from season 1993/1994 to 2010/2011 inclusive (a total of 6624 occurrences). This information is available online at [14].

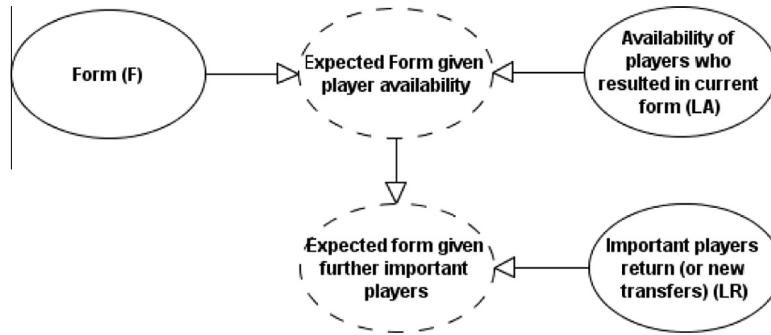


Fig. 4. Level 2 Component (F): measuring expected form ϕ_{LR} . Fig. G.4 presents an example of this component with different scenarios. Dashed nodes indicate latent variables.

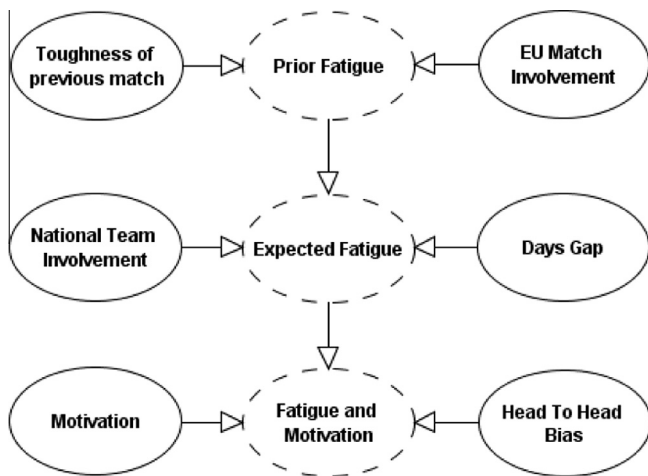


Fig. 5. Component 3 (M): measuring expected fatigue and motivation G. Fig. G.5 presents an example of this component with different scenarios. Dashed nodes indicate latent variables.

approach generates forecasts that are ‘anonymous’ in the sense that historical outcomes are not restricted by the name of the team. For example, given a match between Manchester United and Newcastle United, and assuming their respective S values are $85(S_R = 2)$ and $62(S_R = 7)$, the resulting forecasts will represent: “a team with a probability density function $S(S_R)$ and a maximum likelihood estimation of $85(2)$ plays against a team with a probability density function $S(S_R)$ and a maximum likelihood estimation of $62(7)$ ” instead of: “Man United plays against Newcastle”. Accordingly, a team’s S distribution varies throughout the season, and it is possible for teams to share similar such distributions at certain periods throughout the season.

Fig. 1 illustrates a simplified model topology of the overall Bayesian network model. Fig. G.1 presents the actual outcomes of the Arsenal vs. Liverpool match as forecasted on August 20th 2011 and demonstrates how match forecasts transform on the basis of hierarchical posterior predictive distributions of S beliefs. The observed outcome was A (score was 0–2). Table G.1 provides a brief description of all the model parameters.

The primary differences with the BN model proposed in [7] are:

1. P, which formulates the prior predictive distribution of S , is now measured using a straightforward *Beta – Binomial* approach (which we describe in detail in Section 2.1 below), rather than the complex non-symmetric Bayesian parameter learning approach;
2. Model components which correspond to each of the generic factors have been both decreased in number and simplified in an attempt to reduce model complexity. Specifically:

- (a) The number of variables in component F has decreased from 21 (10 for each team plus one representing discrepancy) to 10 (5 for each team). In particular, instead of providing distinct subjective beliefs about the availability of the *Primary Key Player*, the *Secondary Key Player*, the *Tertiary Key Player*, and the *Remaining First Team Players*, we now introduce a single subjective variable called *Availability of players who resulted in current form*. Further, there is no *Home Form* and *Away Form*, but rather a single variable called *Form* which represents the most recent (and overall) form of a team. This variable is taken into consideration when playing both home and away. As we demonstrate in Section 4.2 below, this modification not only simplified the model, but also resulted in notably increased profitability.
- (b) The two previously proposed model components of *Fatigue* and *Psychological Impact* have been merged into the single component M , and the number of variables (which correspond to each of the two competing teams) has been decreased from a total of 28 to 18 (9 for each team). In particular, instead of requiring indications about the number of first team players rested during the last match in an attempt to measure *Restness*, which is later used to diminish tiredness, we now directly provide this information in a single subjective variable called *Toughness of previous match* (i.e. the subjective indication of toughness will be lower if we already know that some first team players were rested). Further, the beliefs regarding managerial impact, team spirit, motivation, and the expert’s degree of certainty with regards to their subjective indications are now replaced by a single subjective variable called *Motivation* (this follows the same rationale as with the toughness of previous match).
3. In the previous model, the subjective components were used to directly revise the probability distribution of match forecasts in an ordinal manner (i.e. a function was used to skew the predictive distribution of a match forecast towards a home win or an away win based on subjective proximity). Instead of modifying the match forecast directly, we now let the subjective components modify each team’s distribution S , which serve as parents for formulating match forecasts. Thus, match forecasts are now always formulated based on relevant historical data (i.e. no skewness), but given modified distributions of S as a consequence of one or more positive or negative subjective indications. This change does not necessarily reduce model complexity, but rather improves model sophistication and thus, forecasting capability.
4. In the previous model, the values of each of the subjective components (i.e. form, fatigue and psychological impact) were compared between the two teams, and a revision in prediction was only made on the basis of discrepancies between the components (e.g. the team with better form received an

Table 2
Hypothetical betting performance on the basis of profitability between two models.

Match instance	α				β			
	Stake	Return	Profit/loss	Profit rate	Stake	Return	Profit/loss	Profit rate
M_1	£0	£0	–	–	£100	£200	+£100	100%
M_2	£100	£200	+£100	100%	£100	£200	+£100	100%
M_3	£0	£0	–	–	£100	£0	-£100	-100%
M_4	£0	£0	–	–	£100	£200	+£100	100%
M_5	£100	£200	+£100	100%	£100	£200	+£100	100%
Total	£200	£400	+£200	100% ^a	£500	£800	+£300	60% ^a

^a Profit rate based on total stakes.

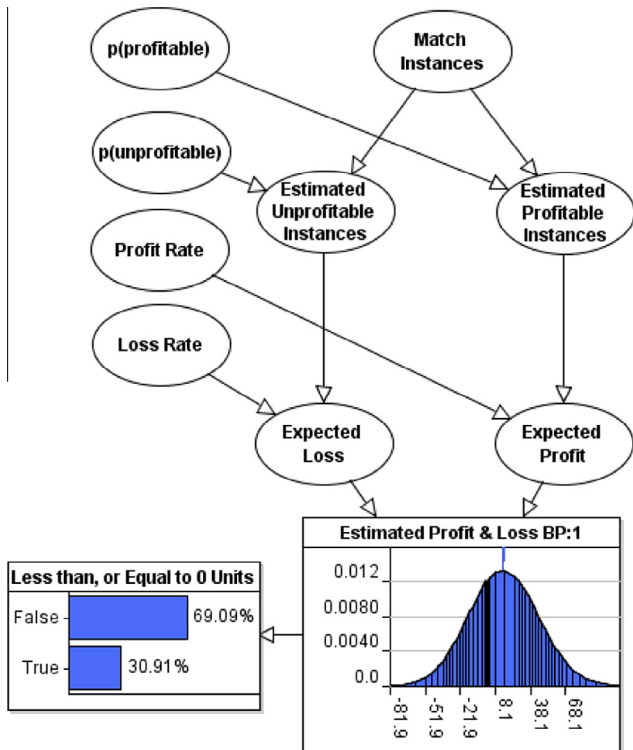


Fig. 6. Bayesian network component for assessing the risks, of accumulating returns below n units, for each of the betting procedures. An example of how the risk of ending with $U \leq 0$ can be measured given BP_1 at discrepancy levels of 0%.

increased probability to win). This implies that match forecasts are not revised if both teams have high (or low) levels of form (i.e. no discrepancy). The problem with this is that, if we assume that high levels of form increase a team's distribution S by 20 points, there is still a difference between the match $S_{HOME} = 60$ versus $S_{AWAY} = 30$, and another $S_{HOME} = 80$ versus $S_{AWAY} = 50$. Consequently, in the new model we do not perform comparisons when it comes to subjective components, but we instead allow each of the components to have a direct impact on each team's distribution S . Again, this change does not necessarily reduce model complexity but rather improves model sophistication.

5. The impact of each model component is now inferred hierarchically; implying that model components now follow a non-linear weighting when revising distribution S (i.e. model components computed first have less impact); in contrast to the previous model, where the three subjective components had identical impact on match forecasts. The hierarchical computation also decreases the time required to calculate posterior probabilities.

2.1. Level 1 component: team performance (P) and inconsistency (I)

At level 1, P is modelled using the *Beta – Binomial* approach. The *Beta* distributions serve as conjugate distributions of the *Binomial* distributions, formulating a compound distribution such that the p parameter of the *Binomial* distribution is being randomly drawn from the *Beta* distribution. In our model, the *Beta* distributions $p(\text{Win})$, $p(\text{Draw})$ and $p(\text{Lose})$ (with hyperparameters α and β priors based on relevant historical data), serve as the p parameters of the *Binomial* distributions *Number of Wins*, *Number of Draws*, and *Number of Loses* respectively.⁵

The posterior *Beta* distributions are then used to estimate team expectations for the residual match instances of the current season. These team expectations allow expert modifications based on subjective beliefs regarding the difficulty of residual opponents (this indication allows the expert to diminish the bias in cases where the results are formulated on the basis of mostly poor/high quality opponents). Observed and expected cumulative match points are then considered for formulating the prior distributions of S . This is the first part (out of two) of level 1. The Bayesian network component P is illustrated in Fig. 2, where:

- the variables $p(\text{Win})$, $p(\text{Draw})$ and $p(\text{Lose})$ are the *Beta* distributions. For example, in the case of $p(\text{Win})$ the hyperparameters are $\sim \text{Beta}(w + 1, d + l + 2)$,⁶ where w is the number of previous season's wins, d is the number of draws, l is the number of losses, and values 1 and 2 are introduced for minimal *Laplacian smoothing* so that we avoid overfitting by ensuring that posterior parameters α and β are positive for all teams;
- the variables *Number of Wins*, *Number of Draws* and *Number of Loses* are $\sim \text{Binomial}(n, p)$. For example in the case of *Number of Wins*, n is the number of matches played during the current season and p is the probability of success for each trial (for this example p is the *Beta* distribution $p(\text{win})$);
- the variable *Expected Residual Points* (π_p) represents the points a team expects to accumulate over the current season's residual match instances and hence, π_p is dependent on the *Number of residual matches* and the posterior *Beta* beliefs of $p(\text{Win})$, $p(\text{Draw})$ and $p(\text{Lose})$;⁷
- the variable *ERP given opponent difficulty* (π_e) is a π_p posterior given the *Difficulty of residual opponents* (ψ), whereby π_e may receive adjustments for up to $\pm 10\%$ based on a 7-level subjective belief, and it is defined as the case function of:

⁵ Effectively a *Dirichlet – multinomial* distribution, where $p(\text{Win}) + p(\text{Draw}) + p(\text{Lose}) = 1$.

⁶ Hyperparameters are provided as node-inputs and are not shown in Fig. 2.

⁷ We do not perform convolution but we instead perform aggregation of averages (which means that the variance might be overestimated) in order to keep the complexity of the model at significantly lower levels.

Table 3
Unit based profit and profitability rates based on BP_1 and BP_2 .

Discrep. levels (%)	Betting Procedure 1 (BP_1)				Betting Procedure 2 (BP_2)			
	Bets/trials	Win rate (%)	P/L (Units)	Profit rate (%)	Bets/trials	Win rate (%)	P/L (Units)	Profit rate (%)
0	379	34.30	15.25	4.02	575	31.83	47.71	8.30
1	359	34.54	17.34	4.83	495	32.53	47.13	9.52
2	316	34.49	15.52	4.91	403	32.75	36.95	9.17
3	272	34.19	5.09	1.87	319	31.97	7.63	2.39
4	227	35.24	13.03	5.74	257	33.85	24.74	9.63
5	193	35.23	11.53	5.97	211	34.12	20.87	9.89
6	168	35.71	17.45	10.39	179	34.64	23.74	13.26
7	144	37.50	8.6	5.97	150	36.67	15.84	10.56
8	129	38.76	15.22	11.80	131	38.17	13.22	10.09
9	107	37.38	-3.67	-3.43	108	37.04	-4.67	-4.32
10	97	39.18	3.31	3.41	97	39.18	3.31	3.41
11	84	35.71	-2.77	-3.30	84	35.71	-2.77	-3.30
12	67	34.33	-6.42	-9.58	67	34.33	-6.42	-9.58
13	53	30.19	-17.02	-32.11	53	30.19	-17.02	-32.11
14	38	34.21	-6.88	-18.11	38	34.21	-6.88	-18.11
15	33	36.36	-6.28	-19.03	33	36.36	-6.28	-19.03

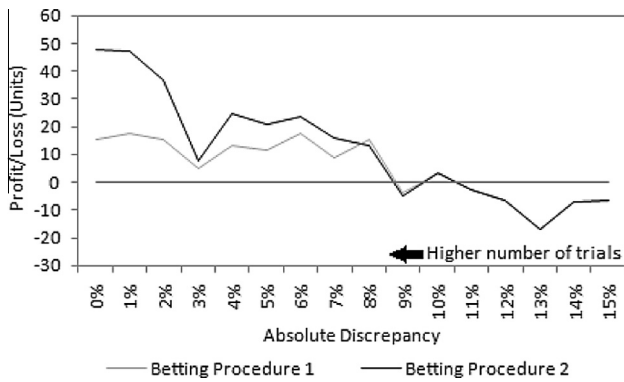


Fig. 7. Unit-based returns based on BP_1 and BP_2 , and according to the specified level of discrepancy.

$$\pi_e = \begin{cases} \min(114, \pi_p \times 1.1), & \pi_p, \psi = \text{Lowest} \\ \min(114, \pi_p \times 1.0666), & \pi_p, \psi = \text{Very Low} \\ \min(114, \pi_p \times 1.0333), & \pi_p, \psi = \text{Low} \\ \min(114, \pi_p), & \pi_p, \psi = \text{Normal} \\ \min(114, \pi_p \times 0.9666), & \pi_p, \psi = \text{High} \\ \min(114, \pi_p \times 0.9333), & \pi_p, \psi = \text{Very High} \\ \min(114, \pi_p \times 0.9), & \pi_p, \psi = \text{Highest} \end{cases}$$

- (e) the variable *Current Points* simply represents the total number of points accumulated over the current season and hence, it is dependent on the relevant *Binomial* observations (see *Table G.1* for details);
- (f) *Team Strength*(S) is then simply the sum of *Current Points* and π_e .

The component inconsistency (I) approximates a team’s inconsistency based on respective league point totals over the five most recent seasons, and the resulting variance is added to the prior predictive distribution of S and together formulate S_{L1} . *Fig. 3* presents a naive parameter learning procedure for approximating a team’s inconsistency, where:

- (a) the variables *Season Y1* to *Y5* are $\sim TNormal(\mu, \sigma^2, 0, 114)$;
- (b) the variable *Inconsistency (Variance)* (V) is a $\sim Uniform(0, 150)$ ⁸ and serves as the input σ^2 for the *TNormal* distributions of (a) above;

- (c) the variable *Overall Performance* is a $\sim Uniform(0, 114)$ and serves as the input μ for the *TNormal* distributions of (a) above.

Moreover, the variable *Confidence in Historical Inconsistency* (C), presented in *Figs. 1* and *G.1*, gives the option to the expert to diminish the additional variance V if the expert feels that the team is not currently as inconsistent as it used to be over the period of the last five seasons. The case function below shows how V , which serves as an input for S_{L1} , can diminish in value based on the subjective indication of C :

$$S_{L1} = \begin{cases} TNormal(S, \frac{V}{3}, 0, 114), & S, C = \text{Low} \\ TNormal(S, \frac{V}{2}, 0, 114), & S, C = \text{Medium} \\ TNormal(S, V, 0, 114), & S, C = \text{High} \end{cases}$$

2.2. Level 2 component: team form (F)

At level 2 posterior predictive distributions of S_{L2} are formulated given S_{L1} and a posterior team-form (Φ), as presented in *Figs. 4* and *G.4*, where Φ is a continuous variable on a scale that goes from 0 to 1. A value close to 0.5 suggests that the team is performing as expected, whereas a higher value indicates that the team is performing better than expected (and vice versa). The expectations are determined by the forecasts generated by this model, and Φ is measured on the basis of the five most recent gameweeks.⁹

The Φ posterior is formulated hierarchically based on the *Availability of players who resulted in current form* (L_A) and the *Important players return* (L_R), where both variables follow ordinal scale distributions with subjective indications as illustrated in *Figs. 4* and *G.4* and the case functions below. The variable *Expected Form given player availability* (Φ_{LA}) is the case function:

$$\Phi_{LA} = \begin{cases} TNormal(\Phi, 0.0001, 0, 1), & \Phi, L_A = \text{VeryHigh} \\ TNormal((\Phi \times 0.8), 0.001, 0, 1), & \Phi, L_A = \text{High} \\ TNormal((\Phi \times 0.6), 0.005, 0, 1), & \Phi, L_A = \text{Medium} \\ TNormal((\Phi \times 0.4), 0.01, 0, 1), & \Phi, L_A = \text{Low} \\ TNormal((\Phi \times 0.2), 0.05, 0, 1), & \Phi, L_A = \text{Very Low} \end{cases}$$

and the variable *Expected form given further important players* (Φ_{LA}) is the case function:

⁸ Upper bound is 150 rather than 114 to account for the limited number of parameters learned.

⁹ A complete EPL season consists of 38 gameweeks.

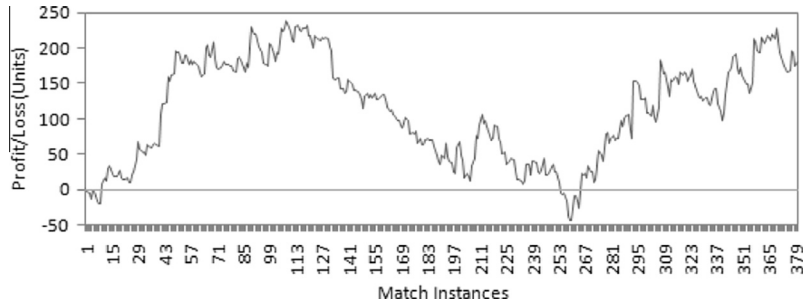


Fig. 8. Cumulative unit-based returns based on BP₃.

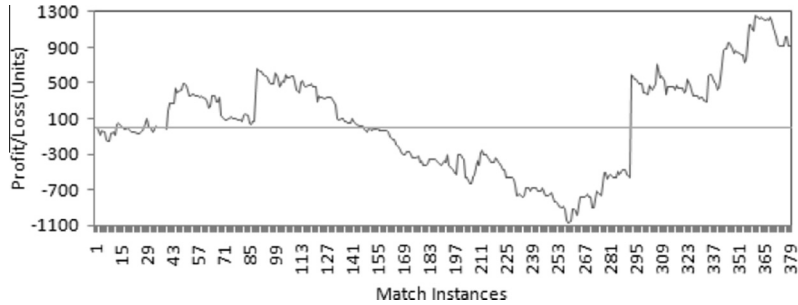


Fig. 9. Cumulative unit-based returns based on BP₄.

Table 4
Risk probability values for the specified concluding returns per betting procedure. Results assume no discrepancy restrictions (set to 0%) for BP₁, BP₂, BP_{5.1}, BP_{5.2}, and an initialised bankroll of 10,000 for the betting procedures of series 5.

BP	Expected profit/loss (less than)									
	U1000 (%)	U500 (%)	U100 (%)	U50 (%)	U0 (%)	-U50 (%)	-U100 (%)	-U500 (%)	-U1000 (%)	
1	100.00	100.00	99.69	87.80	30.91	1.36	0.03	0.00	0.00	
2	100.00	100.00	94.27	53.01	7.61	0.23	0.02	0.00	0.00	
3	99.98	95.13	34.16	25.16	17.53	11.60	7.22	0.08	0.01	
4	53.95	32.70	18.63	17.19	15.76	14.49	13.24	5.95	1.72	
5.1	100.00	81.21	0.00	0.00	0.00	0.00	0.00	0.00	0.00	
5.2	100.00	66.56	0.00	0.00	0.00	0.00	0.00	0.00	0.00	
5.3	97.80	16.32	0.08	0.05	0.02	0.01	0.01	0.00	0.00	
5.4	61.56	31.19	13.20	11.65	10.10	8.86	7.65	2.06	0.27	

$$\Phi_{LR} = \begin{cases} \text{TNormal}(\Phi_{LA}, 0.01, 0, 1), & \Phi_{LA}, L_R = \text{None} \\ \text{TNormal}((\Phi_{LA} + ((1 - \Phi_{LA}) \times 0.1)), 0.01, 0, 1), & \Phi_{LA}, L_R = \text{Low} \\ \text{TNormal}((\Phi_{LA} + ((1 - \Phi_{LA}) \times 0.2)), 0.01, 0, 1), & \Phi_{LA}, L_R = \text{Medium} \\ \text{TNormal}((\Phi_{LA} + ((1 - \Phi_{LA}) \times 0.3)), 0.01, 0, 1), & \Phi_{LA}, L_R = \text{High} \end{cases}$$

2.3. Level 3 component: fatigue and motivation (M)

At level 3 posterior predictive distributions of S_{L3} are formulated given S_{L2} and Φ_{LR} as presented in Fig. 5. A Prior Fatigue(G_p) is first measured given EU match Involvement(E) (which represents team involvement in European tournaments) and Toughness of previous match(T), where E and T follow ordinal scale distributions with subjective indications as illustrated in Figs. 5 and G.5, and the case function of G_p below:

$$G_p = \begin{cases} \text{TNormal}(T, 0.001, 0, 1), & T, E = \text{None} \\ \text{TNormal}((T + (1 - T) \times \frac{1}{6}), 0.001, 0, 1), & T, E = \text{Very Low} \\ \text{TNormal}((T + (1 - T) \times \frac{2}{6}), 0.001, 0, 1), & T, E = \text{Low} \\ \text{TNormal}((T + (1 - T) \times \frac{3}{6}), 0.001, 0, 1), & T, E = \text{Medium} \\ \text{TNormal}((T + (1 - T) \times \frac{4}{6}), 0.001, 0, 1), & T, E = \text{High} \\ \text{TNormal}((T + (1 - T) \times \frac{5}{6}), 0.001, 0, 1), & T, E = \text{Very High} \end{cases}$$

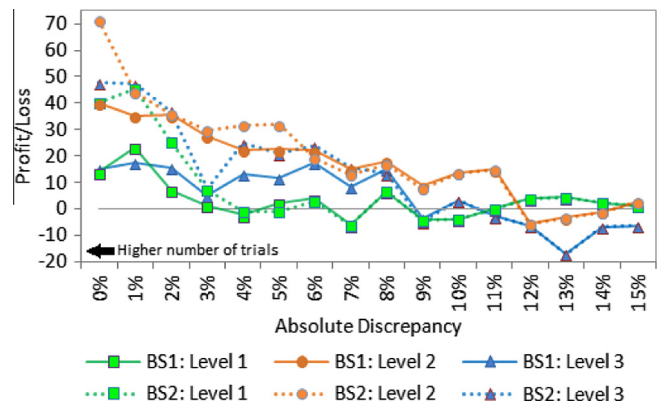


Fig. 10. Cumulative unit-based returns based on BP₁ and BP₂, for component levels 1, 2 and 3.

The Expected Fatigue(G_e) is a posterior G_p value which diminishes on the basis of Days Gap since previous match(δ), and increases with National Team Involvement(λ), where δ and λ are ordinal scale

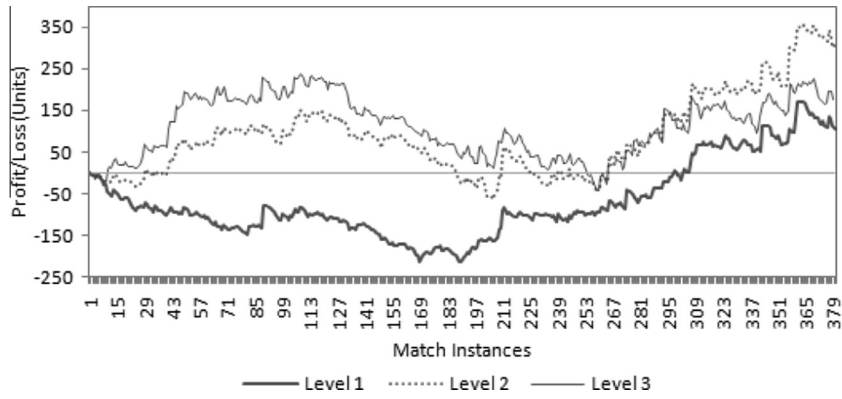


Fig. 11. Cumulative unit-based returns based on BP_3 , for component levels 1, 2 and 3.

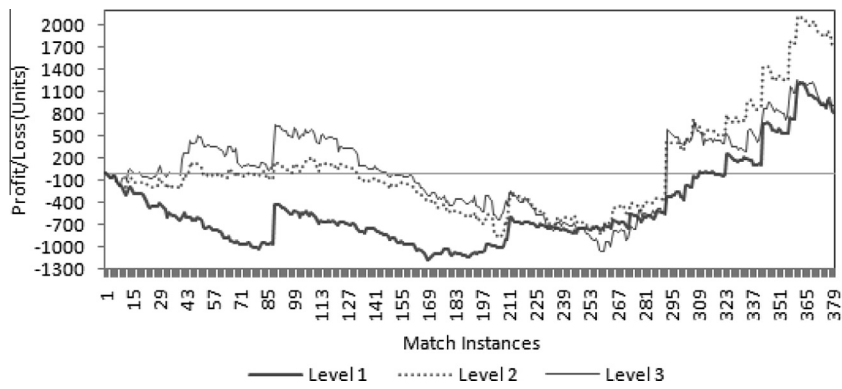


Fig. 12. Cumulative unit-based returns based on BP_4 , for component levels 1, 2 and 3.

Table 5
Team-based returns relative to overall returns for the specified betting procedure.

Rank	Team	Betting procedure								Average
		1 (%)	2 (%)	3 (%)	4 (%)	5.1 (%)	5.2 (%)	5.3 (%)	5.4 (%)	
1	Man City	-28.00	-21.59	-17.96	-36.49	9.88	9.75	8.93	3.65	-8.98
2	Man Utd	-37.57	-14.46	21.83	-24.01	6.35	6.34	8.05	6.47	-3.37
3	Arsenal	111.74	49.49	68.91	59.98	4.82	4.93	7.18	16.71	40.47
4	Tottenham	-25.84	15.97	32.22	8.78	12.14	12.07	12.39	9.77	9.69
5	Newcastle	76.20	19.77	83.19	39.44	10.43	10.33	13.22	19.25	33.98
6	Chelsea	-97.38	-9.16	-108.64	-112.74	11.51	11.80	9.60	3.11	-36.49
7	Everton	-32.39	-12.66	-27.98	-30.82	13.82	13.82	12.75	9.55	-6.74
8	Liverpool	175.87	76.32	192.25	237.84	27.83	27.69	29.59	36.40	100.47
9	Fulham	-25.18	17.84	-10.08	7.17	5.66	5.94	5.30	7.18	1.73
10	West Brom	62.23	-8.55	23.44	31.22	14.67	14.38	14.67	15.96	21.00
11	Swansea	59.54	2.68	-7.67	7.64	7.29	7.09	6.31	6.29	11.15
12	Norwich	-55.93	2.45	-47.79	-32.66	7.72	7.65	6.92	4.51	-13.39
13	Sunderland	-15.61	9.47	-24.50	-24.76	4.52	4.52	4.42	3.06	-4.86
14	Stoke	16.79	36.62	15.39	-12.31	6.88	7.24	7.41	5.75	10.47
15	Wigan	-121.84	4.38	3.66	95.50	9.06	9.22	7.78	8.09	1.98
16	Aston Villa	-70.95	20.29	-25.35	-20.34	7.23	7.73	6.39	4.33	-8.83
17	QPR	128.59	17.80	59.61	91.06	4.88	4.69	6.01	19.33	41.50
18	Bolton	29.70	2.62	5.47	-2.27	7.36	7.25	7.65	9.16	8.37
19	Blackburn	-9.84	-24.90	-33.66	-52.58	11.20	10.95	9.99	3.39	-10.68
20	Wolves	59.87	15.64	-2.34	-29.65	16.73	16.62	15.45	8.03	12.55

distributions with subjective indications as illustrated in Figs. 5 and G.5, and the case function of G_e below:

$$G_e = \begin{cases} \text{TNormal}((G_p - G_p \times \delta), 0.001, 0, 1), & G_p, \delta, \lambda = \text{None} \\ \text{TNormal}(((G_p - G_p \times \delta) + (1 - (G_p - G_p \times \delta)) \times 0.1), 0.001, 0, 1), & G_p, \delta, \lambda = \text{Low} \\ \text{TNormal}(((G_p - G_p \times \delta) + (1 - (G_p - G_p \times \delta)) \times 0.2), 0.001, 0, 1), & G_p, \delta, \lambda = \text{Medium} \\ \text{TNormal}(((G_p - G_p \times \delta) + (1 - (G_p - G_p \times \delta)) \times 0.3), 0.001, 0, 1), & G_p, \delta, \lambda = \text{High} \\ \text{TNormal}(((G_p - G_p \times \delta) + (1 - (G_p - G_p \times \delta)) \times 0.4), 0.001, 0, 1), & G_p, \delta, \lambda = \text{Very High} \end{cases}$$

Table 6
Previous model's profitability based on BP₁ and BP₂ (for season 2010–2011).

Discrep. levels (%)	Betting procedure 1 (BP ₁)				Betting procedure 2 (BP ₂)			
	Bets/trials	Win rate (%)	P/L (Units)	Profit rate (%)	Bets/trials	Win rate (%)	P/L (Units)	Profit rate (%)
0	378	34.66	5.70	1.51	571	31.87	15.55	2.72
1	358	33.52	-1.76	-0.49	485	31.34	-5.55	-1.14
2	325	32.92	-4.79	-1.47	407	31.20	-10.67	-2.62
3	275	33.09	2.85	1.04	324	31.17	-11.19	-3.45
4	225	33.78	11.87	5.28	254	31.89	2.30	0.91
5	169	33.73	14.19	8.40	186	32.80	13.07	7.03
6	131	35.11	17.40	13.28	141	34.75	19.61	13.91
7	107	35.51	12.92	12.07	111	35.14	14.07	12.68
8	84	33.33	8.43	10.04	87	33.33	10.58	12.16
9	71	33.80	11.36	16.00	74	33.78	13.51	18.26
10	52	34.62	10.61	20.40	53	35.85	14.76	27.85
11	41	36.59	14.61	35.63	41	36.59	14.61	35.63
12	25	24.00	-6.95	-27.80	25	24.00	-6.95	-27.80
13	15	26.67	-4.61	-30.73	15	26.67	-4.61	-30.73
14	12	25.00	-3.70	-30.83	12	25.00	-3.70	-30.83
15	10	30.00	-1.70	-17.00	10	30.00	-1.70	-17.00

Finally, G_e is revised into *Fatigue and Motivation* (G) given *Motivation* (κ) and *Head-to-Head Bias* (ω), where κ and ω follow ordinal scale distributions that go from 0 to 1 with subjective indications as illustrated in Figs. 5 and G.5, and the case function of G below:

$$G = \begin{cases} \text{TNormal}(\left(\frac{\kappa+\omega}{2}\right), 0.01, 0, 1), & \kappa, \omega, G_e = \text{Very Rested} \\ \text{TNormal}(\left(\left(\frac{\kappa+\omega}{2}\right) \times 0.9\right), 0.01, 0, 1), & \kappa, \omega, G_e = \text{Rested} \\ \text{TNormal}(\left(\left(\frac{\kappa+\omega}{2}\right) \times 0.8\right), 0.01, 0, 1), & \kappa, \omega, G_e = \text{Normal} \\ \text{TNormal}(\left(\left(\frac{\kappa+\omega}{2}\right) \times 0.7\right), 0.01, 0, 1), & \kappa, \omega, G_e = \text{Tired} \\ \text{TNormal}(\left(\left(\frac{\kappa+\omega}{2}\right) \times 0.6\right), 0.01, 0, 1), & \kappa, \omega, G_e = \text{Very Tired} \end{cases}$$

3. Forecast performance based on profitability and risk

In this section we describe how the forecasting capability of the model was assessed on the basis of profitability and relevant risks involved. Profitability is measured on the basis of a set of predetermined betting procedures. For market odds we have considered the odds with the highest payoff as recorded by [14] for the matches of the EPL season 2011/2012. The number of bookmaking firms considered for recording maximums ranged from 26 to 49 per match instance.¹⁰

Naturally, the performance of a football forecast model is determined by its ability to generate profit against market odds. However, many researchers also consider (or solely focus) on various scoring rules for this purpose in an attempt to determine the accuracy of the forecasts against the observed results [8,32,20,16,24,16,15,22,19,21]. Forecast assessments based on scoring rules have been heavily criticised because different rules may provide different conclusions about the forecasting capability of football forecast models [4]. Furthermore, in financial domains researchers have already demonstrated a weak relationship between various accuracy and profit measures [25], whereas [40] suggested that it might be best to combine accuracy and profit measures for a more informative picture.

In this paper we are interested in the profitability of the model relative to market odds. For this to happen, market odds have to be sufficiently less accurate (or inefficient) relative to those generated by our model so that the bookmakers' profit margin, where pres-

ent, can be overcome. The bookmakers' profit margin, sometimes also called 'over-round', refers to the margin by which the sum of published market probabilities of the total outcomes exceeds 1. For example, if the true (i.e. the initially measured) probabilities for a match instance are $p(\text{Win}) = 0.50$, $p(\text{Draw}) = 0.25$ and $p(\text{Lose}) = 0.25$, a bookmaker's published probabilities might be $p(\text{Win}) = 0.52$, $p(\text{Draw}) = 0.26$ and $p(\text{Lose}) = 0.26$ (which result in lower odds for payoff) and hence, the sum of published probabilities exceeds 1. The bookmaker's profit margin here is simply $(p(\text{Win}) + p(\text{Draw}) + p(\text{Lose})) - 1$; which in this case would be 4%.

Since profitability is not only dependent on the forecasting capability of a model relative to market odds but also on the specified betting methodology, we have introduced an array of such betting procedures. For each procedure, we introduce sensible modifications relative to the standard betting strategy that was proposed and considered by the vast majority of the previous relevant published papers, whereby a bet is placed when expectations exceed a predetermined level [30,8,32,9,17,15,19,21,7].

3.1. Defining profitability

We measure the profitability on the basis of the quantity of profit (or net profit which is stated as unit-based returns), rather than on the basis of percentage returns relative to respective stakes. The example below illustrates the rationale behind our preference.

Example: Suppose we have two football forecast models α and β . We want to compare their performance on the basis of profitability given the set of five match instances $\{M_1, M_2, M_3, M_4, M_5\}$. Table 2 presents a hypothetical betting performance between the two models over those match instances.

After considering the five match instances we observe the following results¹¹:

- Model α suggested two bets and both were successful (100% winning rate), returning a net profit of £200 which represents a profit rate of 100% relative to total stakes.
- Model β suggested five bets and four of them were successful (80% winning rate), returning a net profit of £ 300 which represents a profit rate of 60% relative to total stakes.

¹⁰ Betfair odds are not considered within the dataset since Betfair is a betting exchange company whereby published odds constantly fluctuate. These odds are normally the best possible odds (i.e. with the highest payoff) a bettor can find online. However, unlike traditional bookmakers Betfair will deduct a fixed % from your winnings which ranges from 2% to 6% depending on membership status [2].

¹¹ For simplification we assume identical stakes (£100) and odds for payoff (*evens*; or 2.00 in decimal form).

An evaluation based on the percentage profit rates would have erroneously considered model β as being inferior at picking winners than model α . But, such an evaluation fails to consider the possibility that model α might have failed to discover potential advantages against the market for all of the match instances. The reality is that model β managed to simulate riskier bets that reduced the percentage rates of winning and profit, but increased net profit due to the larger number of successful bets.

We have to choose which model is best to follow; model α with a higher winning rate on bets and a higher profit rate between stakes and returns, or model β with a higher (33.33%) net profit? If the ultimate aim is to make money, then every bettor would have preferred model β over model α for betting against the market. Therefore, we suggest that a bettor should be increasing net profit rather than establishing good winning percentage rates, and for this to happen a bettor is expected to consider all of his advantages presented at every match instance rather than choosing the 'best' of his advantages that occasionally arise.

Consequently, in this paper we measure profitability on unit-based returns (net profit) over n match instances (in our case $n = 380$, the total number of matches played in the EPL season of 2011/2012). The betting procedures are defined in the following section.

3.2. Defining the betting procedures

We define the following set of betting procedures for evaluating the profitability of the model against the market:

1. (BP_1): For each match instance, place a fixed bet equal to a single unit on the outcome with the highest absolute percentage discrepancy, where the model predicts the higher probability, if and only if the discrepancy is $\geq n\%$ (where n is an integer $0 \leq n \leq 15$);
2. (BP_2): For each match instance, place a fixed bet equal to a single unit on every outcome the model predicts with higher probability, if and only if the absolute discrepancy is $\geq n\%$;
3. (BP_3): For each match instance, place a bet equal to U units for each outcome the model predicts with higher probability, where the stake of the bet is a real number equal to the absolute discrepancy percentage between outcomes multiplied by U (e.g. if an absolute discrepancy of 4.45% and 1.17% is observed for outcomes H and D respectively while $U = 1$, then bets of £4.45 and £1.17 are simulated for a home win and a draw respectively);
4. (BP_4): For each match instance, place a bet equal to U units for each outcome the model predicts with higher probability, where the stake of the bet is a real number equal to the relative discrepancy percentage between outcomes multiplied by U (e.g. if a relative discrepancy of 4.45% and 1.17% is observed for outcomes H and D respectively while $U = 1$, then bets of £4.45 and £1.17 are simulated for a home win and a draw respectively);
5. ($BP_{5,1}$, $BP_{5,2}$, $BP_{5,3}$, $BP_{5,4}$): These apply only to match instances where arbitrage¹² opportunities are discovered. Repeat 1, 2, 3 and 4 but substitute the betting procedure with arbitrage bets whereby the total amount of the three bets is equal to the bank-

roll available at that time (a bankroll specification is required prior to initialising the betting simulation, and tests are performed for different bankroll values).

If a betting procedure A indicates higher profitability than another B over a fixed number of match instances, it does not necessarily suggest that we should always choose A over B . This is true if we are also interested in the risks involved and the level of uncertainty over the posterior predicted distribution of unit-based returns (i.e. the magnitude of potential losses and winnings as well as the probability associated with such events). Accordingly, we have constructed a simple Bayesian network component (Fig. 6) that measures the risk of ending with less than, or equal to, a specified number of units over a specified number of match instances. Fig. 6 illustrates, as an example, the risk of ending with $U \leq 0$ after bets are simulated (given BP_1 at discrepancy levels of 0%) on the 380 match instances. This assumes relevant model performances as demonstrated in Section 4 below. In particular,

- (a) the variable *Match Instances* represents the number of match instances over which the risk is measured;
- (b) the variables p (*profitable*) and p (*unprofitable*) are *Beta* distributions with α and β hyperparameters representing the probability to profit (and not to profit) for each match instance simulated;
- (c) the variables *Estimated Unprofitable Instances* and *Estimated Profitable Instances* are *Binomial* distributions with n number of trials equal to (a) above, where input p is the respective *Beta* distribution of (b) above;
- (d) the variables *Profit Rate* and *Loss Rate* are averaged values associated with observed profit and loss for respective match instances;
- (e) the variables *Expected Loss* and *Expected Profit* are posterior predictive density functions which represent the overall loss/profit given (c) and (d) above;
- (f) the variable *Estimated Profit & Loss* is the summary probability density function given (e);
- (g) the variable *Less than, or Equal to 0 Units* is the probability of ending at, or below the specified value of U given (f) above.

4. Results and discussion

In this section we demonstrate and discuss the resulting performance of the model. In Section 4.1 we demonstrate the profitability of the model along with the relevant risks involved with each of the betting procedures; in Section 4.2 we evaluate the effectiveness of the model components based on the transitions of profitability at each hierarchical component level; in Section 4.3 we provide evidence of market inefficiency based on specific football teams; finally, in Section 4.4 we compare the performance of the model against the model presented in [7].

4.1. Model performance

Table 3 presents the amount of bets simulated and unit-based returns (along with the frequency rates of successful bets and profit rate relative to stakes for procedures BP_1 and BP_2) at the specified discrepancy levels. Fig. 7 illustrates a summary comparison between the two betting procedures. In general, under both procedures the model appears to be profitable at discrepancy levels up to 10%, but unprofitable thereafter. In particular, for BP_1 the profitability appears to be consistent up to that point, with the highest returns of U17.45 and U17.34 observed at discrepancy levels of 6% and 1% respectively. In contrast, BP_2 generated maximum returns that are substantially higher relative to BP_1 ; returns of

¹² "An arbitrage opportunity is simply an opportunity whereby profit is guaranteed on the basis of a negative profit margin which results by combining the odds published by the various bookmaking firms. In particular, arbitrage opportunities depend on two factors: (a) the divergence in outcome probabilities between bookmaking firms and (b) the profit margin by each bookmaker. Negative profit margin is simply a scenario where a set of HDA probabilities is found (for a single match instance) in which the sum of the probabilities within that set is < 1 . Hence, profit for the bettor can be guaranteed if the bets are placed such that the return is identical whatever the outcome." [6].

U47.71 and U47.13 at discrepancy levels of 0% and 1% respectively. Figs. A.1 and A.2 compare the cumulative returns over the season between the two betting procedures; the results show that BP_2 consistently generates higher returns than BP_1 throughout the period and at almost every discrepancy level.

At discrepancy levels of $\geq 11\%$ BP_2 essentially mimics the betting simulation of BP_1 since it becomes unlikely for probabilities of paired match instances (model and market) to encompass more than one outcome at such high discrepancy levels. At discrepancy levels of $\geq 10\%$ the model appears to be unprofitable, with betting trials in the range of 33 and 84. However, it would not be safe to formulate conclusions on the basis of model performances at such high discrepancy levels. We explain why next.

For BP_1 and BP_2 , it is important to note that we are much more confident about results generated at lower discrepancy levels, since at those levels the number of bets simulated is sufficiently high for us to formulate safe conclusions. As the discrepancy levels increase, the number of betting trials inevitably decreases. Yet, at higher discrepancy levels we actually require *more* betting trials to formulate conclusions that are as safe as those at the lower levels. To understand why, assume that we have simulated 50 bets at discrepancy levels of $\geq 11\%$. Among the 50 there will be lots of instances of the following:

- (a) Team A plays B and A is a strong favourite, but not as strong as the bookies think. Consequently, the bookies offer a probability of just 5% that team B wins. The model, however rates the probability as 17% and so we bet on team B to win (if we consider discrepancy levels of $\geq 12\%$). If the model is 'correct' we would still only win about once every eight match instances of this 'type'. Therefore, 50 trials is not a sufficiently high number to formulate conclusions. For instance, Fig. 7 shows that an additional successful bet at decimal odds of approximately 15.00 would lead to profitable returns at almost all of the discrepancy levels above 10%, which demonstrates the high level of uncertainty.
- (b) Team A plays B and A is a strong favourite, but stronger than the bookies think. The bookies offer a probability of 70% that team A wins, while the model rates the probability as 82%. So we bet on team A to win (again, if we consider discrepancy levels of $\geq 12\%$). If the model is 'correct' we would win about four times for every five bets simulated. In this case, most bets win. However, when they periodically occur the returns from winning match instances are too small to compensate for the high uncertainty generated on the basis of numerous instances of (a).

It should also be noted that the occurrence rate of the above two cases is likely to be affected by the well known phenomenon of the favourite longshot-bias observed by the markets.¹³

Figs. 8 and 9 demonstrate the cumulative unit-based returns given BP_3 and BP_4 respectively. In both cases, considerably higher returns are generated relative to BP_1 and BP_2 . In particular, the concluding balance of BP_3 at match instance 380 is U180.34, whereas for BP_4 it is U922.97. Since BP_4 is a replicative version of BP_3 (with the difference that stakes generated are based on the relative, rather than the absolute, discrepancy of model to market

probabilities), it is normal for BP_4 to generate cumulative returns that are excessive versions of those of BP_3 . The cumulative distributions in Figs. 8 and 9 show that BP_3 experienced a maximum loss of -U43.65 (81.63% less relative to its maximum profit of U237.57), whereas BP_4 experienced a maximum loss of -U1066.33 (14.54% less relative to its maximum profit of U1247.86). Further, BP_4 remained at a state of loss for a longer period throughout the season, whereas BP_3 remained at a state of loss for only a period of 11 match instances (out of 380). Table 4 presents the risk probability values for ending up with less than, or equal to, the specified concluding profit/loss balances according to the specified betting procedure, and Fig. B.1 presents the respective predicted probability density risk distributions.

4.1.1. Arbitrage opportunities and risk assessment

There are various ways to reduce our exposure to risk. In our case, a straightforward solution would be to take advantage of existing arbitrage opportunities and replace the betting procedure with arbitrage bets when such risk free match instances are exposed. In fact, 70 match instances (out of the 380) allowed for risk free returns for the season under study, where arbitrage betting guaranteed an average profit of 0.57% per such match instance with minimum and maximum risk free returns at 0.03% and 1.94% respectively. Figs. C.1, C.2, C.3 and C.4 demonstrate how the profit rate converges relative to an initialised bankroll on the basis of $BP_{5,1}$, $BP_{5,2}$, $BP_{5,3}$, and $BP_{5,4}$ (as described in Section 3.2). Table 4 and Fig. B.1 demonstrate the reduction in risk and uncertainty, when taking advantage of arbitrage instances, relative to the respective procedures of BP_1 , BP_2 , BP_3 , and BP_4 which do not take advantage of such opportunities. As expected, due to the relatively high number of arbitrage instances the profitability is heavily dependent on the initialised bankroll. When an arbitrage opportunity is discovered the bet is equal to the value of the bankroll at that specific time. Bankrolls with sufficiently high initialised values (i.e. ≥ 1000 or $\geq 10,000$ in this case) eventually overshadow the predictive performance of the model since generated returns converge towards the arbitrage profit rate.

4.2. Effectiveness of model components

Figs. 10–12 demonstrate the transitions of profitability at component levels 1, 2 and 3 given BP_1 , BP_2 , BP_3 and BP_4 . We observe that the model component at level 2 (team form) generates profitability that is substantially superior to that of level 1, for all of the betting procedures. However, profitability is reduced at level 3 (team fatigue and motivation). We have therefore analysed the sub-parameters of that component in an attempt to investigate how they have negatively affected the performance of the model relative to market odds. Figs. D.1, D.2, D.3 and D.4 demonstrate the profitability of the model over procedures BP_1 , BP_2 , BP_3 and BP_4 when:

- (a) we only consider match instances with evidence of *fatigue* (but no evidence of *motivation*);
- (b) we only consider match instances with evidence of *motivation* (but no evidence of *fatigue*);
- (c) we only consider match instances with evidence of both *fatigue* and *motivation*;
- (d) we only consider match instances where neither evidence of *fatigue* nor evidence of *motivation* exist.

Assuming that we rank profitability-based performances from 1 to 4 (1 being best), the results suggest that evidence of *fatigue* provided the worse overall performance with resulting ranks of 3, 4, 4 and 4 under procedures BP_1 , BP_2 , BP_3 and BP_4 respectively.

¹³ The phenomenon whereby bettors have a preference in backing risky outcomes and hence, bookmakers offer more-than-fair odds to 'safe' outcomes, and less-than-fair odds to 'risky' outcomes. This phenomenon is not only observed in football but also in many different markets [1,31,38,33,35,34,41,39,18,23,6]. Various theories exist, such as risk-loving behaviour, on why people are willing to bet on such uncertain propositions [37,36].

This suggests that we have, most likely, overestimated the negative impact of *fatigue* for a team (i.e. the number of days gap since last competing match, the toughness of previous match, the involvement in European competitions, and player participation with their national team). On the other hand, *motivation* (whereby the quality of the input is predominantly dependent on the expert) provided performances with resulting ranks of 4, 1, 3 and 1 under the four respective betting procedures, and signs of improvement (relative to test (d)) in forecasting capability are observed only under two of the four betting procedures.

4.3. Team-based market inefficiency

The results reported in this section add further evidence of market inefficiency to an already extensive list, particularly in the presence of regular predetermined biases, arbitrage opportunities, as well as conflicting daily adjustments in published odds between firms [6]. We also considered a *team-based* profitability assessment (see Table 5), where the percentage values represent the returns U of a team relative to the returns over all teams based on the specified betting procedure.¹⁴

Our results demonstrate notable differences in profitability for five out of the twenty teams. In particular, for match instances involving Liverpool, QPR, Arsenal and Newcastle our model generated notably higher returns relative to the overall team, whereas for match instances involving Chelsea our model generated notably lower returns. Fig. E.1 illustrates the team-based explicit returns throughout the season against market odds for the above five teams. Results show that:

- (a) market odds overestimated the performances of Liverpool at a consistent rate, and particularly over the final third of the season (during which Liverpool accumulated only 10 points during their last 10 matches). This allowed our model to generate profitable returns during the specified period;
- (b) as in (a), the same applies to Arsenal but to a lower extent. This allowed our model to generate profitable returns during the specified period;
- (c) market odds underestimated the performances of Newcastle at a consistent rate, and particularly over the first half of the season. It is important to note that Newcastle finished at position 5 with 65 points after being promoted to the EPL only a season earlier. This allowed our model to generate profitable returns during the specified period;
- (d) we do not consider that market odds underestimated performances of QPR at the absence of consistency and high uncertainty in returns; profit was generated due to a pair of match instances with excessive returns;
- (e) our model overestimated the performances of Chelsea, particularly over the first two thirds of the season, at a consistent rate. This is highly likely to be due to Chelsea's erratic performances under a new manager who was eventually sacked during that period. This led our model to generate unprofitable returns during the specified period. The returns over the final third of the season, during which Chelsea provided more consistent performances under a new manager, appear to be evened.

¹⁴ If for the specified betting procedure a team generates returns A which are equal to the returns B generated by all of the teams (overall), then team A is 100% related to set B .

4.4. Performance comparison against the previously published BN model

Figs. F.1, F.2 and F.3 compare the unit-based cumulative returns over a period of 380 match instances (but for different seasons¹⁵) between the two models. The results show that the new model generates superior returns under all of the betting procedures.¹⁶ In particular, for BP_1 and BP_2 the new model generated increased net-profit of 33.67% and 210.98% respectively. An interesting distinction between the two models (according to the first two betting procedures) is that the previous model provides higher profit rates but lower net-profit due to the significantly lower number of bets simulated (as discussed in Section 3.1, and Tables 3 and 6 verify this behaviour). Further, for scenarios BP_3 and BP_4 the new model generates respective net-profit that is 158.43% and 49.68% higher relative to respective returns from the previous model.

5. Concluding remarks

We have presented a Bayesian network (BN) model for forecasting football match outcomes that not only simplifies a previously published BN model, but also provides improved forecasting capability. The model considers both objective and subjective information for prediction. The subjective information is important for prediction but is not captured in historical data. The model was used to generate the match forecasts for the EPL season 2011/2012, and forecasts were published online [29] prior to the start of each match.

For assessing the forecast capability of our model, we have introduced an array of betting procedures. These are variants of a standard betting methodology previously considered for assessing profitability by relevant published football forecast studies. A unit-based profitability assessment over all betting procedures demonstrates that:

- (a) at level 2 (team form) the model component provided inferred match forecasts that were substantially superior to those generated at level 1 (which were solely based on historical performances);
- (b) at level 3 (team fatigue and motivation) the model component failed to provide inferred match forecasts that were superior to those generated at level 2. This resulted in concluding match forecasts with inferior profitability relative to that of level 2, but still superior relative to that of level 1;
- (c) a sub-component evaluation at level 3 revealed that we have overestimated the negative impact introduced by evidence of *fatigue*, and this should serve as a lesson-learned for relevant future models;
- (d) despite the consequences of (b), the concluding profitability of our model was even superior to that generated by the previous successful and profitable model under all of the betting procedures;
- (e) the predictive probability density distributions of unit-based returns showed that a bettor's exposure to risk increases together with the substantial profitable returns that BP_3 , and BP_4 provide over BP_1 and BP_2 . However, we showed that one way a bettor may reduce his exposure to risk is by exploiting arbitrage opportunities which occur relatively frequently (70 out of the 380 match instances);

¹⁵ We compare the forecasting capability between the two models relative to market odds, where the old version was assessed over the EPL season 2010–2011, and the new version (presented in this paper) over the EPL season 2011–2012.

¹⁶ Following the discussion in Section 4.1, we have ignored the scenarios whereby the discrepancy levels of BP_1 and BP_2 are set to $\geq 11\%$.

(f) a team-based profitability assessment revealed further market inefficiencies (to the already extensive list) whereby published odds are consistently biased towards the *trade-mark* rather than the *performance* of a team.

Evidently, the results of our study are critically dependent on the knowledge of the expert. Given that the subjective model inputs were provided by a member of the research team (who is a football fan but definitely not an expert of the EPL), it suggests that (a) subjective inputs can improve the forecasting capability of a model even if they are not submitted by a genuine expert who is a professional for the specified domain, and (b) if the model were to be used by genuine experts we would expect that the more informed expert inputs would lead to posterior beliefs that are even higher in both precision and confidence.

The results of this paper have demonstrated a number of benefits of using Bayesian networks: in particular they enable us to incorporate crucial subjective information easily and enhance our understanding of uncertainty and our exposure to the relevant risks involved.

Acknowledgements

We acknowledge the financial support by the Engineering and Physical Sciences Research Council (EPSRC) for funding this research and Agena Ltd for software support.

Appendix A. Cumulative Returns based on BP_1 and BP_2

Figs. A.1 and A.2.

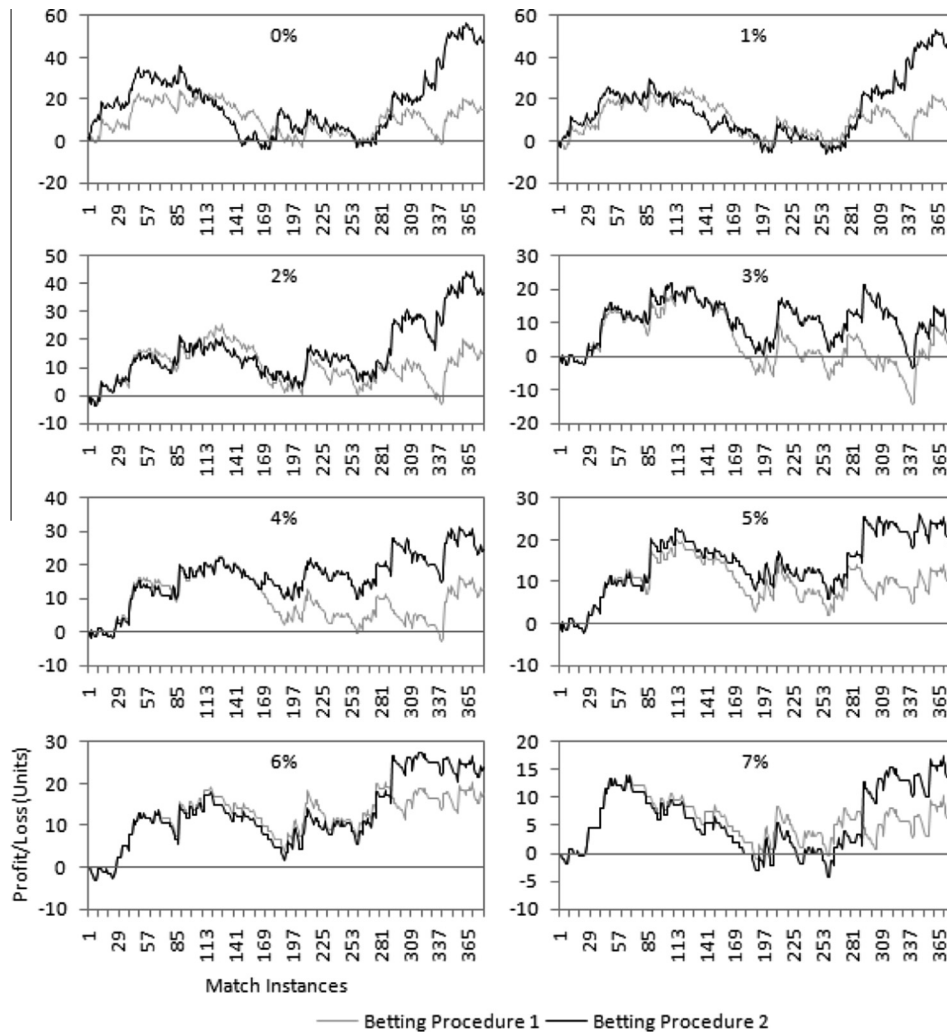


Fig. A.1. Cumulative unit-based returns based on BP_1 and BP_2 according to the specified discrepancy level.

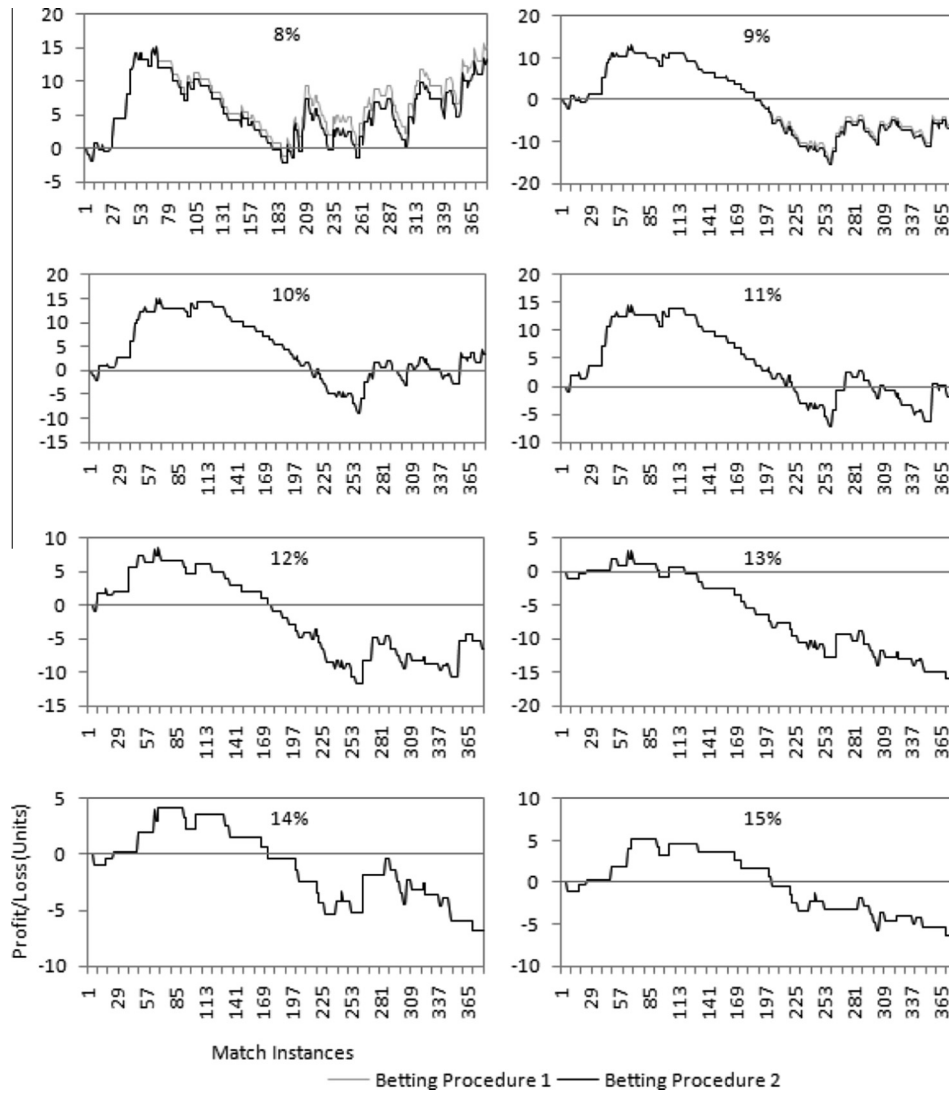


Fig. A.2. Cumulative unit-based returns based on BP_1 and BP_2 according to the specified discrepancy level.

Appendix B. Risk Assessment of Profit and Loss based on the specified betting procedure

Fig. B.1.

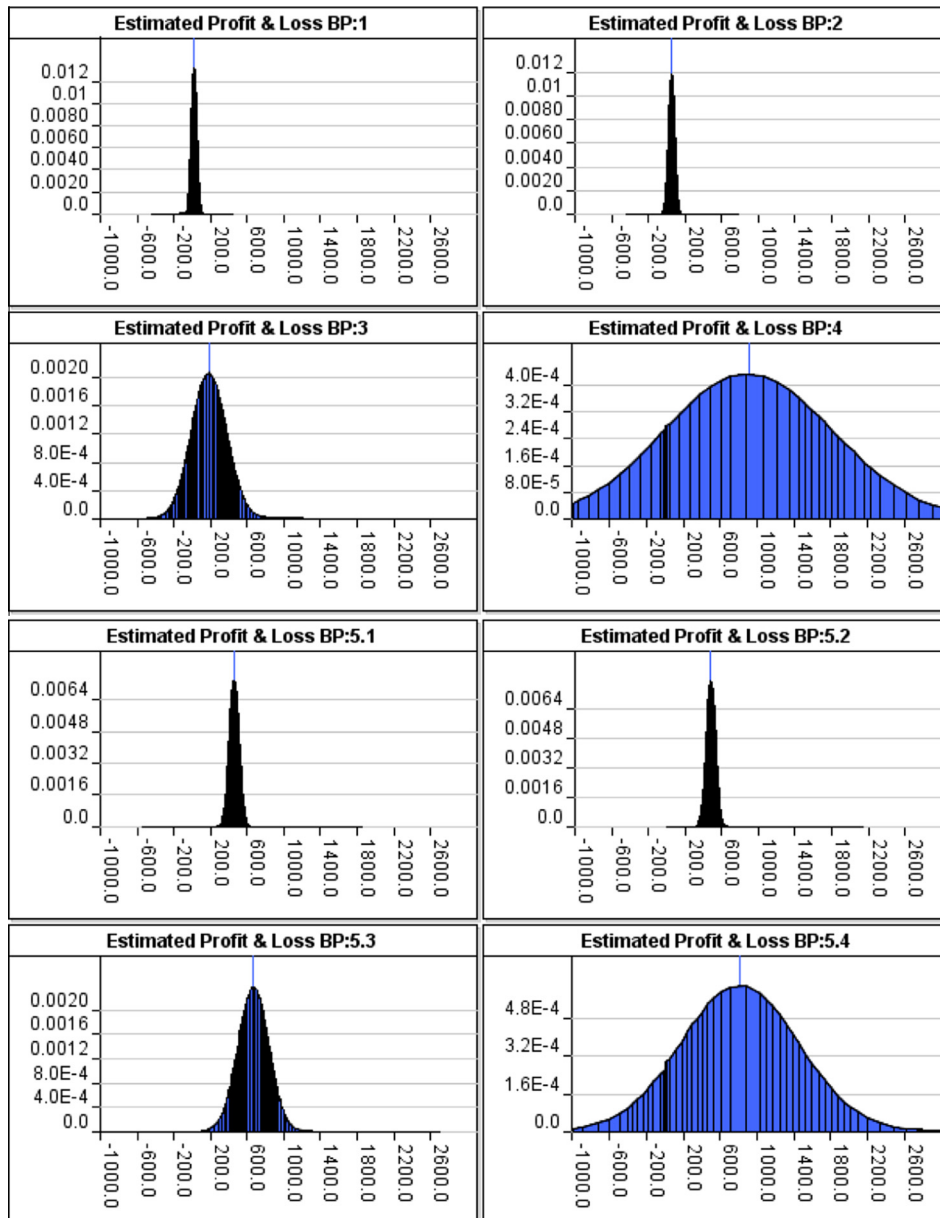


Fig. B.1. Risk assessment of expected returns for each of the betting procedures.

Appendix C. Model performance when considering arbitrage opportunities

Figs. C.1,C.2,C.3 and C.4.

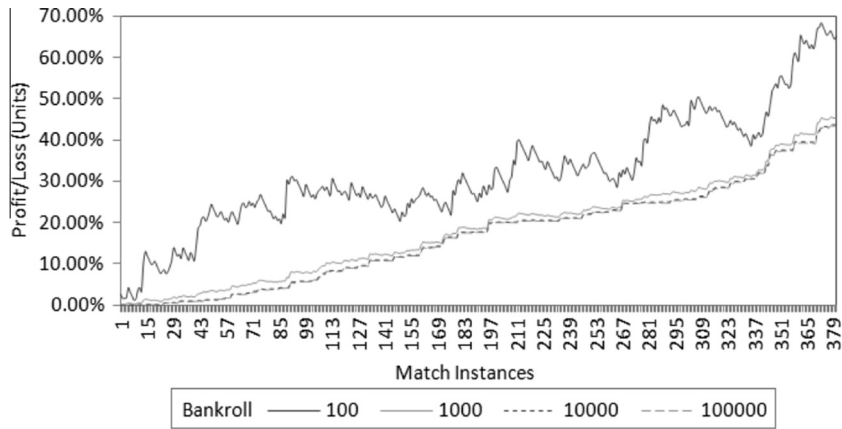


Fig. C.1. Cumulative unit-based returns based on $BP_{5,1}$ assuming no discrepancy restrictions (set to 0%) and according to the specified bankrolls prior to initialising the betting simulation.

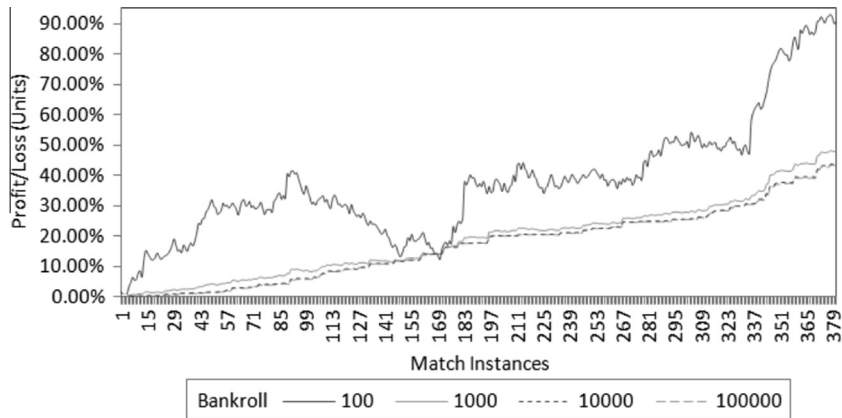


Fig. C.2. Cumulative unit-based returns based on $BP_{5,2}$ assuming no discrepancy restrictions (set to 0%) and according to the specified bankrolls prior to initialising the betting simulation.

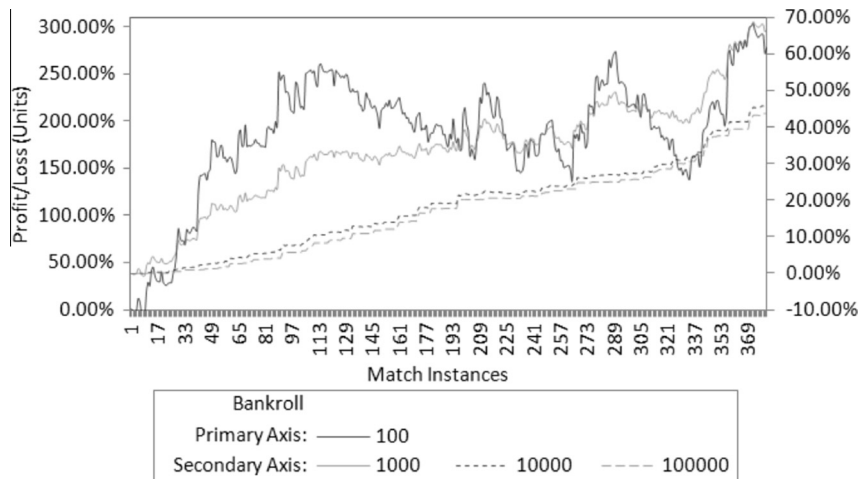


Fig. C.3. Cumulative unit-based returns based on $BP_{5,3}$ and according to the specified bankrolls prior to initialising the betting simulation.

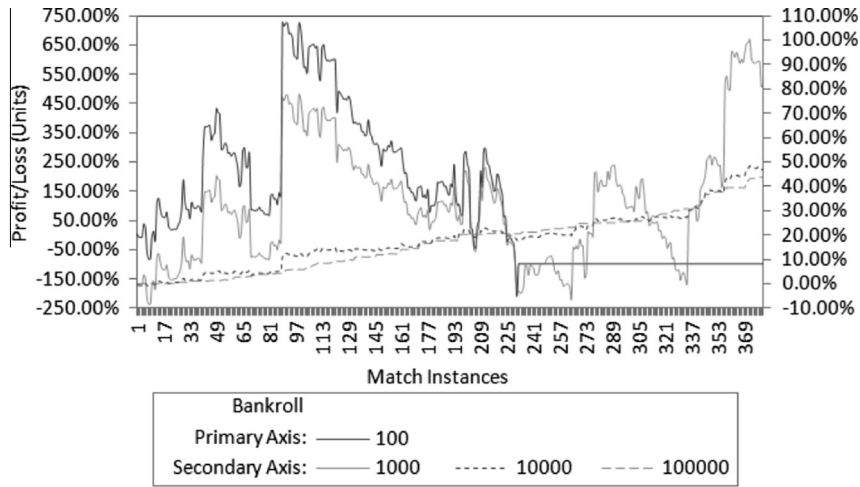


Fig. C.4. Cumulative unit-based returns based on $BP_{5,4}$ and according to the specified bankrolls prior to initialising the betting simulation.

Appendix D. Performance based on parameters of component level 3

Figs. D.1,D.2,D.3 and D.4.

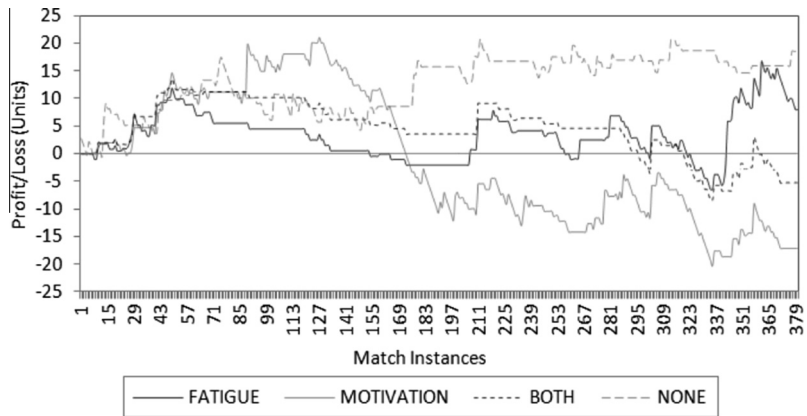


Fig. D.1. Cumulative unit-based returns based on BP_1 for match instances with the specified evidence.

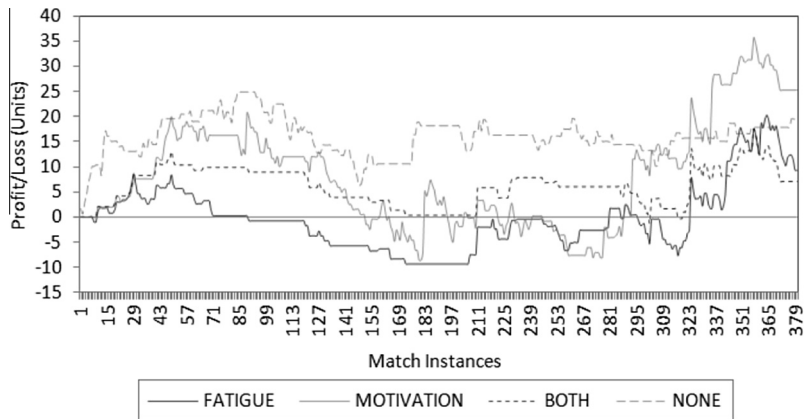


Fig. D.2. Cumulative unit-based returns based on BP_2 for match instances with the specified evidence.

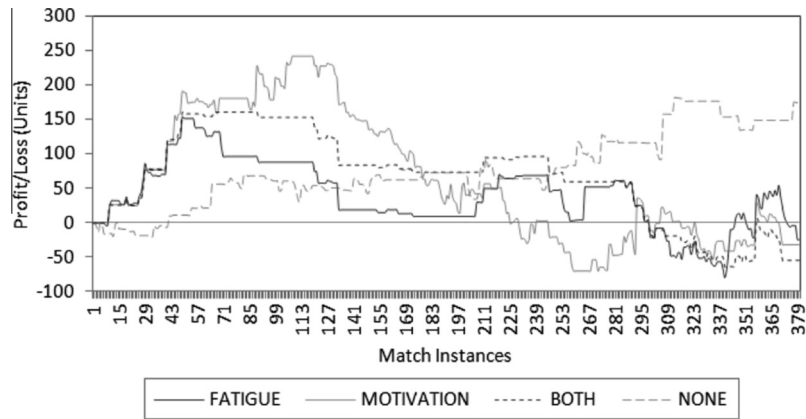


Fig. D.3. Cumulative unit-based returns based on BP_3 for match instances with the specified evidence.

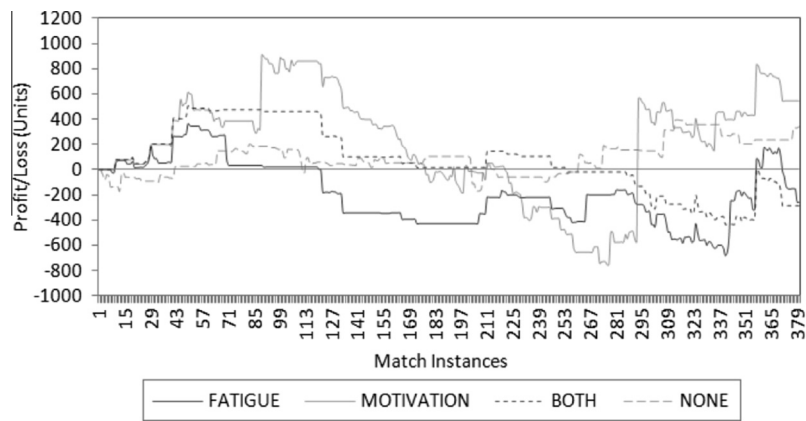


Fig. D.4. Cumulative unit-based returns based on BP_4 for match instances with the specified evidence.

Appendix E. Team-based efficiency

Fig. E.1.

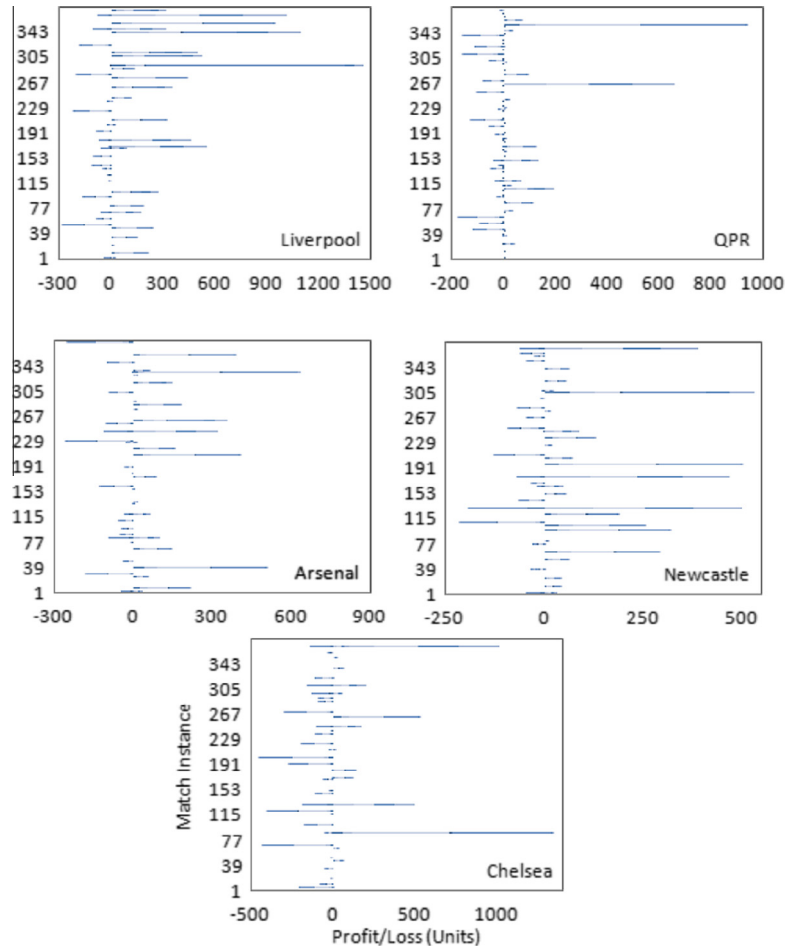


Fig. E.1. Team-based explicit returns against market odds throughout the EPL season.

Appendix F. Unit-based performance relative to the old model

Figs. F.1,F.2 and F.3.

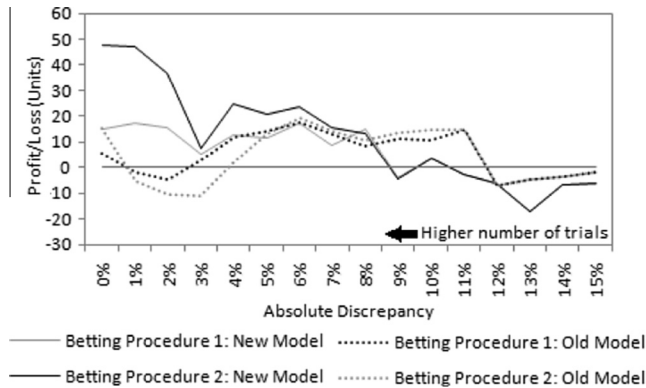


Fig. F.1. Cumulative unit-based returns based on BP_1 and BP_2 ; a comparison between the new and the old model.

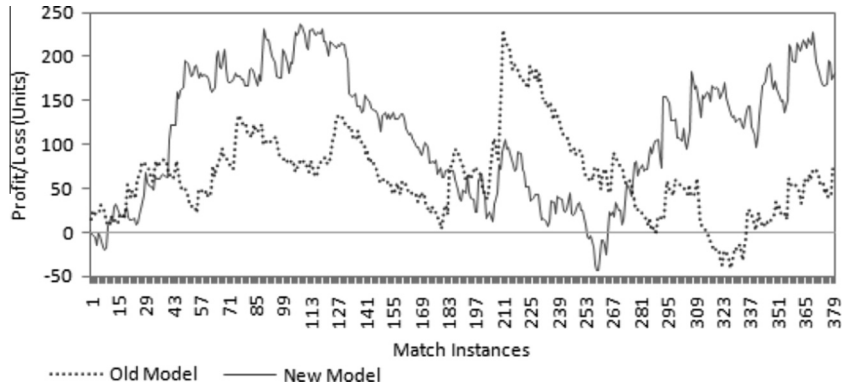


Fig. F.2. Cumulative unit-based returns based on BP_3 ; a comparison between the new and the old model.

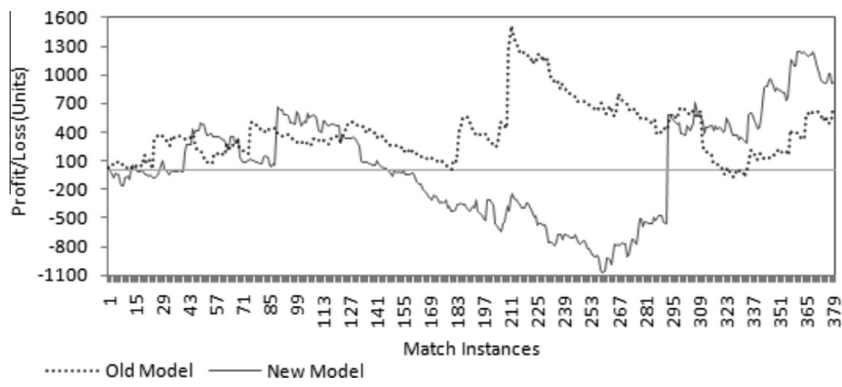


Fig. F.3. Cumulative unit-based returns based on BP_4 ; a comparison between the new and the old model.

Appendix G. Description of model variables and actual examples of the BN model

Table G.1 and Figs. G.1,G.2,G.3,G.4 and G.5.

Table G.1
Brief description of model variables.

Model component	Variable (node) name	Variable type	Observable/latent	Definition	Comments
Level 1_p	Number of Wins	Integer Interval (~Binomial (n, p))	Observable	$\sim \text{Binomial} \left(\begin{matrix} \text{NumberOfMatchesPlayed} \\ p(\text{Win}) \end{matrix} \right)$	Used for inferring $p(\text{Win})$. Same applies to “Number of Draws” and “Number of Loses”
Level 1_p	Number of matches played	Integer interval (Arithmetic)	Definitional	Serves as hyperparameter n for the variables: number of wins, draws, loses, and as a hyperparameter for number of residual matches	Represents the summation of wins, draws and loses. Similarly, the definition of “Number of residual matches” is 38 minus “Number of matches played”
Level 1_p	Current Points	Integer interval (Arithmetic)	Definitional	$\min \left(114, 3 \times \text{NumberOfWins} + \text{NumberOfDraws} \right)$	
Level 1_p	p (Win)	Continuous Interval $\sim \text{Beta} (\alpha, \beta)$	Latent	$\sim \text{Beta} \left(\begin{matrix} 1 + \text{NumberOfWins} \\ 1 + 38 - \text{NumberOfWins} \end{matrix} \right)$	Assumes prior $\sim \text{Beta} (1, 1)$. Same applies to “ p (Draw)” and “ p (Lose)”
Level 1_p	Expected Residual Points	Continuous (Arithmetic)	Latent	$\min \left(114, \text{NumberOfResidualMatches} \times (3 \times p(\text{Win}) + p(\text{Draw})) \right)$	
Level 1_p	Difficulty of residual opponents	Continuous Interval (Ranked)	Observable	7 ordered states from “Lowest” to “Highest”	Represents subjective indications
Level 1_p	ERP given opponent difficulty	Continuous Interval (Arithmetic)	Latent	π_e as defined in Section 2.1	
Level 1_p	Team Strength (S) L	Continuous Interval (Arithmetic)	Latent	$\min (114, \text{ERPgivenOpponentDifficulty} + \text{CurrentPoints})$	

(continued on next page)

Table G.1 (continued)

Model component	Variable (node) name	Variable type	Observable/latent	Definition	Comments
Level 1 _i	Inconsistency (Variance)	Continuous Interval (~Uniform (a, b))	Latent	~Uniform (0, 150)	
Level 1 _i	Overall Performance (mean points)	Continuous Interval (~Uniform (a, b))	Latent	~Uniform (0, 114)	
Level 1 _i	Season y1	Integer Interval (~TNormal (μ, σ ² , a, b))	Observable	~TNormal (OverallPerformance, Inconsistency.001, 0, 1)	The same applies to Seasons y2 to y5
Level 2	Form (Φ)	Continuous Interval (~TNormal (μ, σ ² , a, b))	Observable	~TNormal (Φ, 0.001, 0, 1)	Φ is measured outside of the BN (see Section 2.2)
Level 2	Availability of players who resulted in current form (LA)	Continuous Interval (Ranked)	Observable	5 Ordered states from “Very Low” to “Very High”	Represents subjective indications
Level 2	Important players return (or new transfers) (LR)	Continuous Interval (Ranked)	Observable	4 Ordered states from “None” to “High”	Represents subjective indications
Level 2	Expected Form given player availability	Continuous Interval (~TNormal (μ, σ ² , a, b))	Latent	Φ _{LA} as defined in Section 2.2	
Level 2	Expected form given further important players	Continuous Interval (~TNormal (μ, σ ² , a, b))	Latent	Φ _{LR} as defined in Section 2.2	
Level 3	Toughness of previous match	Continuous Interval (Ranked)	Observable	5 Ordered states from “Very Low” to “Very High”	Represents subjective indications
Level 3	EU Match Involvement	Continuous Interval (Ranked)	Observable	6 Ordered states from “None” to “Very High”	Represents subjective indications
Level 3	National Team Involvement	Continuous Interval (Ranked)	Observable	5 Ordered states from “None” to “Very High”	Represents subjective indications
Level 3	Days Gap	Continuous Interval (~TNormal (μ, σ ² , a, b))	Observable	5 Ordered states from “1–2” to “6+”	Represents subjective indications
Level 3	Motivation	Continuous Interval (Ranked)	Observable	5 ordered states from “Very Low” to “Very High”	Represents subjective indications
Level 3	Head To Head Bias	Continuous Interval (Ranked)	Observable	5 states: “HT Advantage” and “AT Advantage”	Represents subjective indications
Level 3	Prior Fatigue	Continuous Interval (Ranked)	Latent	G _p as defined in Section 2.3	
Level 3	Expected Fatigue	Continuous Interval (Ranked)	Latent	G _e as defined in Section 2.3	
Level 3	Fatigue and Motivation	Continuous Interval (~TNormal (μ, σ ² , a, b))	Latent	G as defined in Section 2.3	
Topology	Confidence in historical inconsistency			3 ordered states from “Low” to “High”	Represents subjective indications
Topology	Team Strength (S) L1	Continuous Interval (~TNormal (μ, σ ² , a, b))	Latent	S _{L1} as defined in Section 2.1	
Topology	Team Strength (S) L2	Continuous Interval (~TNormal (μ, σ ² , a, b))	Latent	μ = if (Φ < 0.5) then: S _{L1} + ((114 – S _{L1}) × (0.5 – Φ)), else: S _{L1} – (S _{L1} × (Φ – 0.5)), σ ² = 1 + ABS(Φ – 0.5) × 10, a = 0, b = 114	The same applies to: “Team Strength (S) L3”, where S _{L1} is replaced by S _{L2} , and Φ is replaced by G

Table G.1 (continued)

Model component	Variable (node) name	Variable type	Observable/latent	Definition	Comments
Topology	Ranked Quality (Level 1)	Integer Interval (~TNormal (μ, σ^2, a, b))	Latent	$\mu = \text{if } (S_{L1} > 89) \text{ then:}$ 1, else: if ($S_{L1} < 20$) then: 14, else: $15 - \left(\frac{S_{L1} - 19}{5}\right)$ $\sigma^2 = 0.01, a = 0, b = 114$	The same applies to: "Ranked Quality (Level 2)" and "Ranked Quality (Level 3)", where S_{L1} is replaced by S_{L2} and S_{L3} respectively
Topology	Level 1 Forecast	Labelled	Latent	Estimated given historical database (i.e. results of match instances which correspond to the two S_{L1} parent nodes)	The same applies to: "Level 2 Forecast" and "Final Forecast", whereby S_{L1} is replaced by S_{L2} and S_{L3} respectively

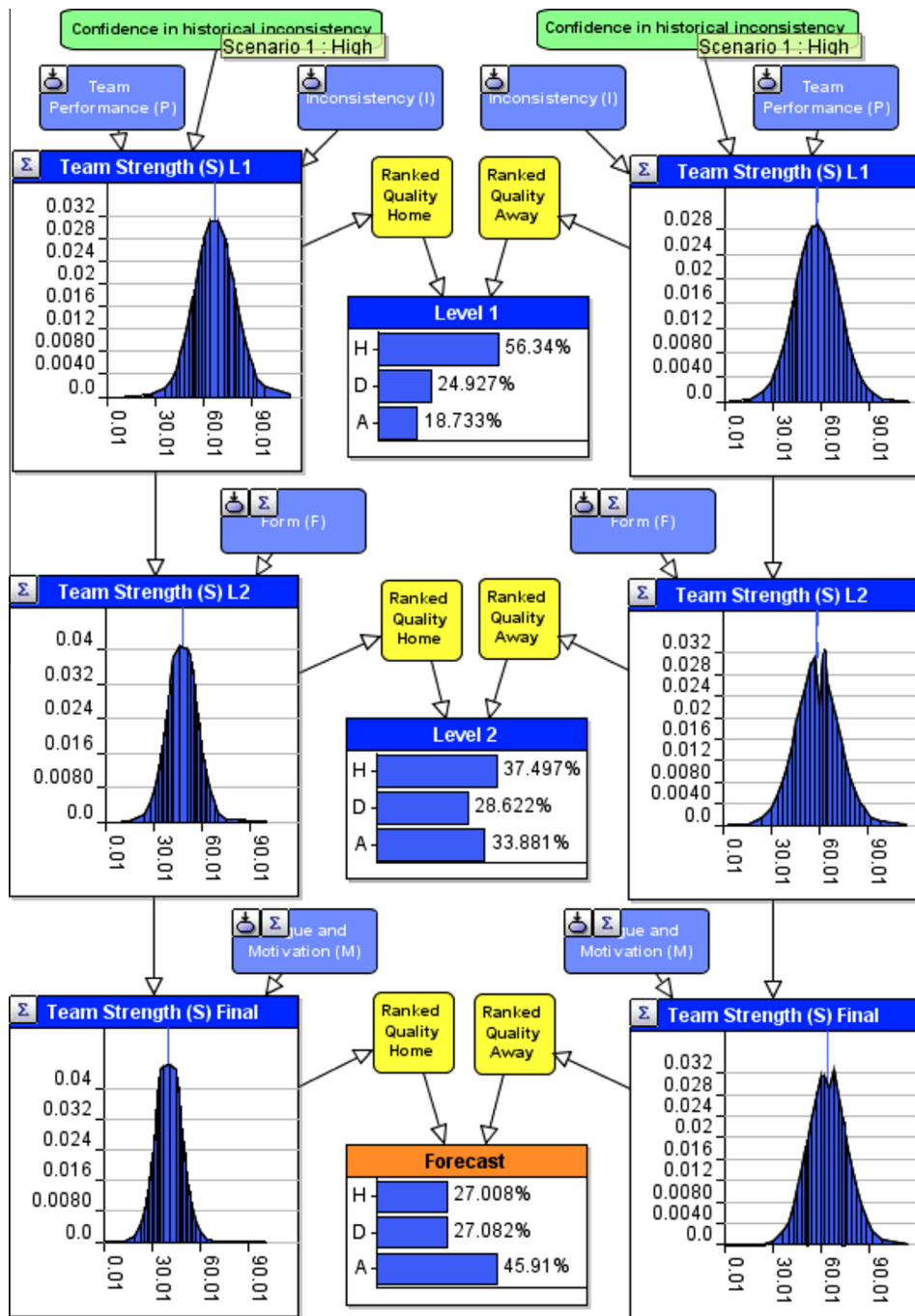


Fig. G.1. A simplified representation of the overall Bayesian network model. An example based on the actual scenarios of the Arsenal vs. Liverpool EPL match, August 20th 2011. The observed outcome was A (0–2).

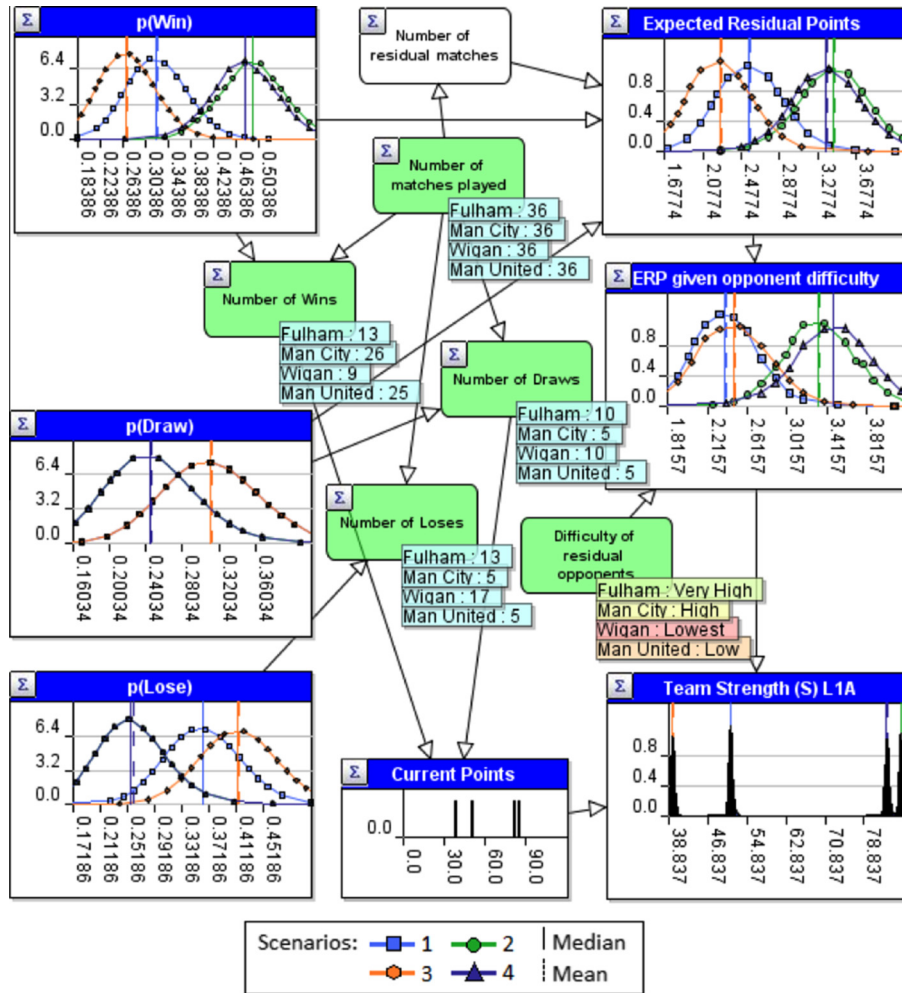


Fig. G.2. Level 1 Component (P): formulating S prior. An example with four actual scenarios based on Fulham, Man City, Wigan, and Man United data, as retrieved at gameweek 37 during season 2011/2012.

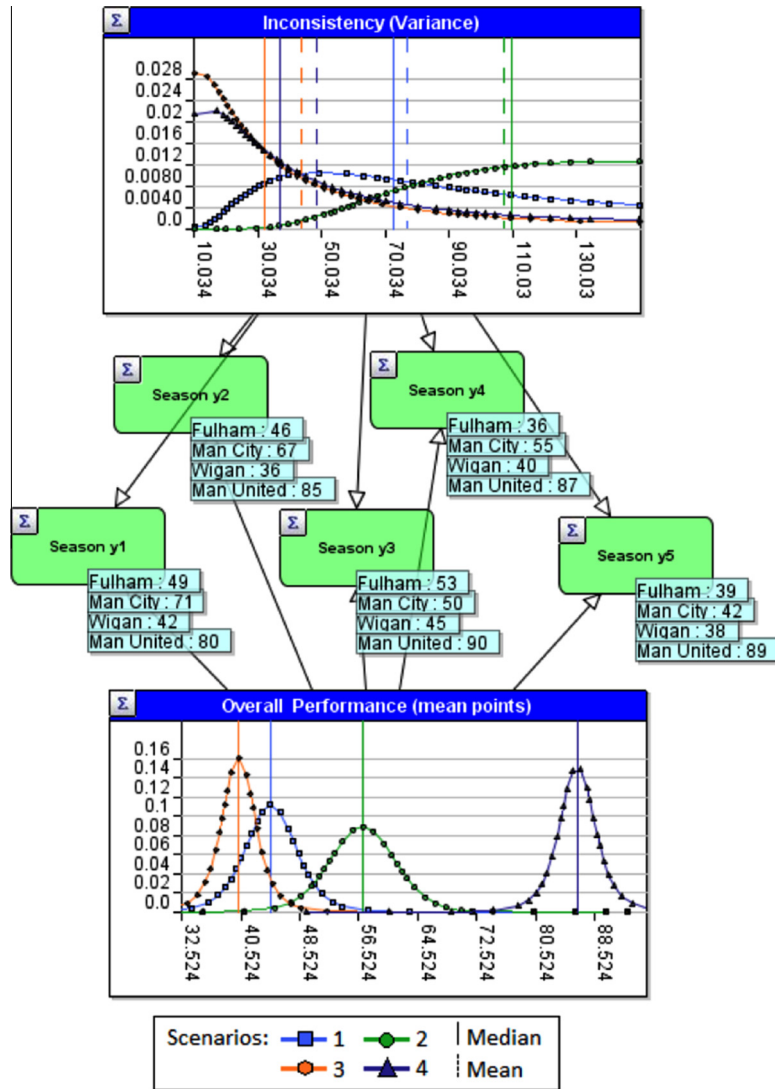


Fig. G.3. Level 1 Component (I): measuring a team's historical inconsistency (V) based on league point totals of the five most recent seasons. An example with four actual scenarios based on Fulham, Man City, Wigan and Man United data for the five seasons preceding EPL 2011/2012.

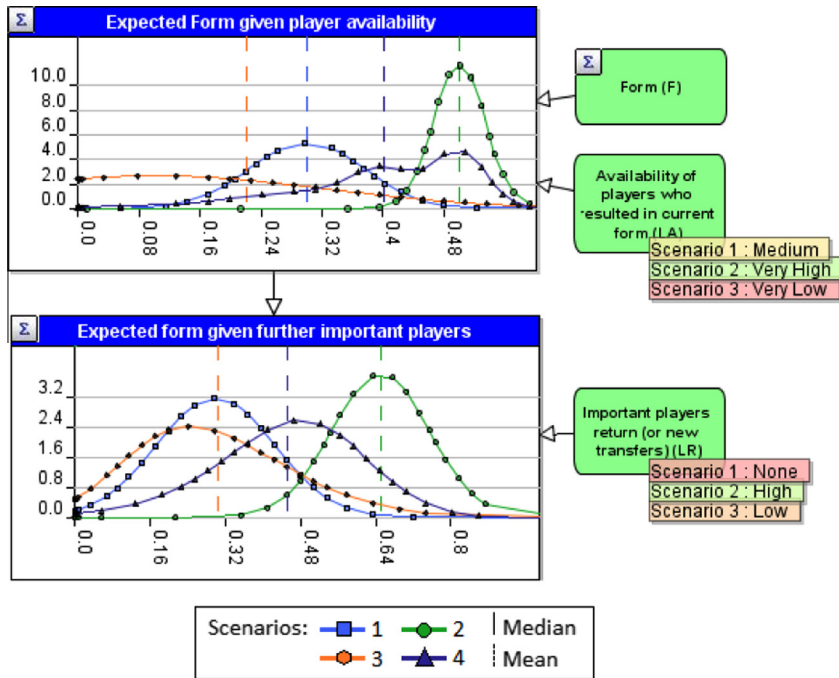


Fig. G.4. Level 2 Component (F): measuring team form. An example with four scenarios (scenario 4 represents uncertain inputs whereby values follow predetermined subjective prior probabilities).

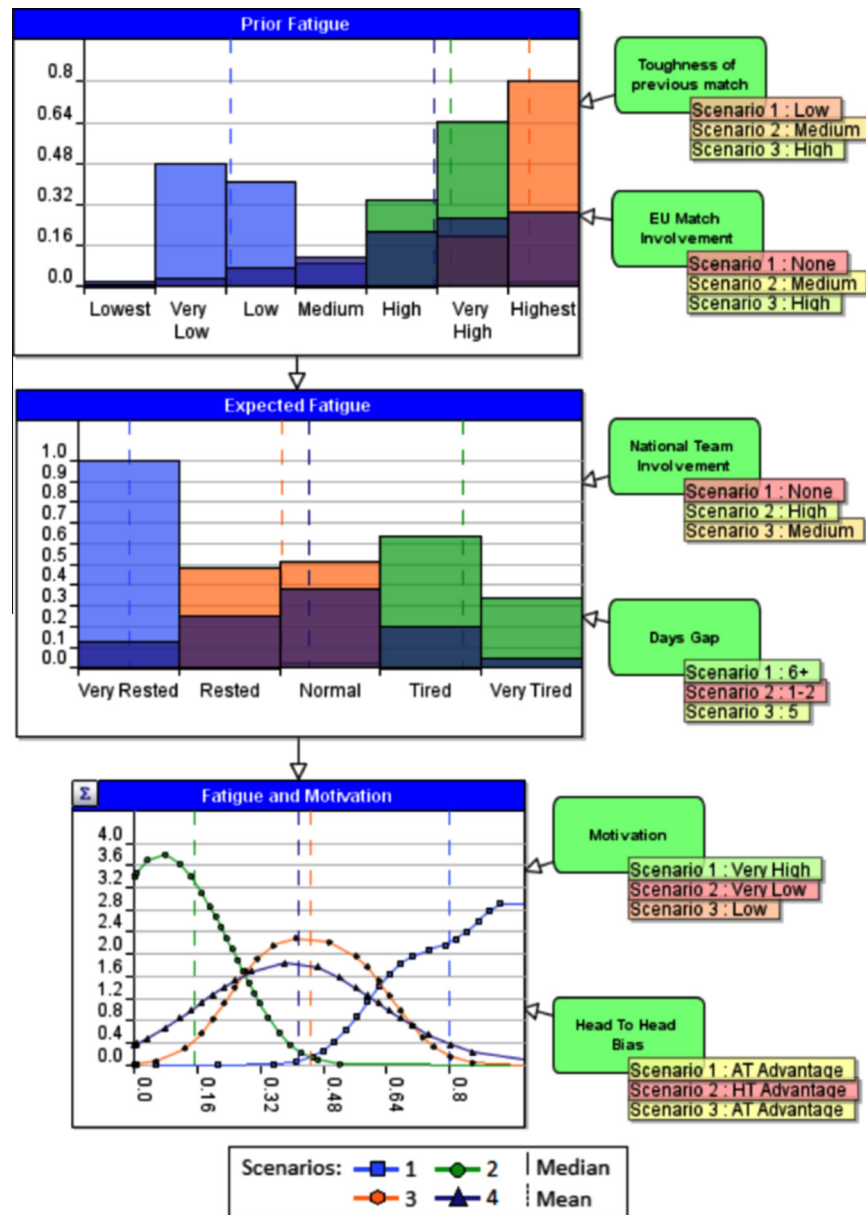


Fig. G.5. Component 3 (M): measuring fatigue and motivation. An example with four scenarios (scenario 4 represents uncertain inputs whereby values follow predetermined subjective prior probabilities).

References

- [1] M. Ali, Probability and utility estimates for racetrack bettors, *Journal of Political Economy* 85 (1977) 803–815.
- [2] Betfair, About Us; How does Betfair work? 2000. <<http://help.betfair.com/contents/itemId/j65767197/index.en.html>> (retrieved 22.11.11).
- [3] M. Cain, D. Law, D. Peel, The favourite-longshot bias and market efficiency in UK football betting, *Scottish Journal of Political Economy* 47 (2000) 25–36.
- [4] A.C. Constantinou, N.E. Fenton, Solving the problem of inadequate scoring rules for assessing probabilistic football forecast models, *Journal of Quantitative Analysis in Sports* 8 (1) (2012) Article 1.
- [5] A.C. Constantinou, N.E. Fenton, Determining the level of ability of football teams by dynamic ratings based on the relative discrepancies in scores between adversaries, *Journal of Quantitative Analysis in Sports* 0 (0) (2013) 1–14.
- [6] A.C. Constantinou, N.E. Fenton, Profiting from arbitrage and odds biases of the European football gambling market, under review. <<http://constantinou.info/downloads/papers/evidenceofinefficiency.pdf>>.
- [7] A.C. Constantinou, N.E. Fenton, M. Neil, pi-football: a Bayesian network model for forecasting Association Football match outcomes, *Knowledge-Based Systems* 36 (2012) 322–339.
- [8] M. Dixon, S. Coles, Modelling association football scores and inefficiencies in the football betting market, *Applied Statistics* 46 (1997) 265–280.
- [9] M. Dixon, P. Pope, The value of statistical forecasts in the UK association football betting market, *International Journal of Forecasting* 20 (2004) 697–711.
- [10] E.G. Dunning, R.E. Joseph A. Maguire, *The Sports Process: A Comparative and Developmental Approach*, Human Kinetics, Champaign, 1993. p. 129.
- [11] E. Dunning, *Sport Matters: Sociological Studies of Sport, Violence and Civilisation*, Routledge, London, 1999.
- [12] A.E. Elo, *The Rating of Chess Players, Past and Present*, Arco Publishing, New York, 1978.
- [13] N.E. Fenton, M. Neil, J.G. Caballero, Using Ranked nodes to model qualitative judgements in Bayesian Networks, *IEEE TKDE* 19 (10) (2007) 1420–1432.
- [14] Football-Data, Football-Data.co.uk. Retrieved May 18, 2012, from Football Results, Statistics & Soccer Betting Odds Data, 2012. <<http://www.football-data.co.uk/englandm.php>>.
- [15] D. Forrest, J. Goddard, R. Simmons, Odds-setters as forecasters: the case of English football, *International Journal of Forecasting* 21 (2005) 551–564.
- [16] J. Goddard, Regression models for forecasting goals and match results in association football, *International Journal of Forecasting* 21 (2005) 331–340.
- [17] J. Goddard, I. Asimakopoulos, Forecasting football results and the efficiency of fixed-odds betting, *Journal of Forecasting* 23 (2004) 51–66.
- [18] J. Golec, M. Tamarkin, Bettors love skewness, not risk, at the horse track, *Journal of Political Economy* 106 (1998) 205–225.
- [19] I. Graham, H. Stott, Predicting bookmaker odds and efficiency for UK football, *Applied Economics* 40 (2008) 99–109.

- [20] N. Hirotsu, M. Wright, An evaluation of characteristics of teams in association football by using a Markov process model, *The Statistician* 4 (2003) 591–602.
- [21] L.M. Hvattum, H. Arntzen, Using ELO ratings for match result prediction in association football, *International Journal of Forecasting* 26 (2010) 460–470.
- [22] A. Joseph, N. Fenton, M. Neil, Predicting football results using Bayesian nets and other machine learning techniques, *Knowledge-Based Systems* 7 (2006) 544–553.
- [23] B. Jullien, B. Salanie, Estimating preferences under risk: the case of racetrack bettors, *Journal of Political Economy* 108 (2000) 503–530.
- [24] D. Karlis, I. Ntzoufras, Analysis of sports data by using bivariate Poisson models, *The Statistician* 3 (2003) 381–393.
- [25] G. Leitch, J.E. Tanner, Economic forecast evaluation: profits versus the conventional error measures, *American Economic Association* (1991) 580–590.
- [26] M.J. Maher, Modelling association football scores, *Statistica Neerlandica* 36 (1982) 109–118.
- [27] F.O. Mueller, R.C. Cantu, S.P. Camp, Catastrophic Injuries in High School and College Sports, *Human Kinetics*, Champaign, 1996. p. 57.
- [28] M. Neil, D. Marquez, N.E. Fenton, Improved reliability modeling using Bayesian networks and dynamic discretization, *Reliability Engineering & System Safety* 95 (4) (2010) 412–425.
- [29] pi-football, Probabilistic Intelligence in Football, Retrieved July 26, 2012, from Forecasts for season 2010/11, 2010. <http://www.pi-football.com/past_predictions.aspx>.
- [30] P. Pope, D. Peel, Information, prices and efficiency in a fixed-odds betting market, *Economica* 56 (1989) 323–341.
- [31] R.E. Quandt, Betting and equilibrium, *Quarterly Journal of Economics* 101 (1986) 201–207.
- [32] H. Rue, O. Salvesen, Prediction and retrospective analysis of soccer matches in a league, *The Statistician* 3 (2000) 339–418.
- [33] H. Shin, Optimal betting odds against insider traders, *Economic Journal* 101 (1991) 1179–1185.
- [34] H. Shin, Measuring the incidence of insider trading in a market for state-contingent claims, *Economic Journal* 103 (1993) 1141–1153.
- [35] R.E. Shin, Prices of state contingent claims with insider traders, and the favourite longshot bias, *Economic Journal* 102 (1992) 426–435.
- [36] E. Snowberg, Explaining the favorite-long shot bias: is it risk-love or misperceptions?, *Journal of Political Economy* 118 (2010) 4, 723–74
- [37] R.S. Sobel, S.T. Raines, An examination of the empirical derivatives of the favourite-longshot bias in racetrack betting, *Applied Economics* 35 (2003) 371–385.
- [38] R. Thaler, W. Ziemba, Parimutuel betting markets: racetracks and lotteries, *Journal of Economic Perspectives* 2 (1988) 161–174.
- [39] L. Vaughn Williams, D. Paton, Why is there a favourite-longshot bias in British racetrack betting markets?, *Economic Journal* 107 (1997) 150–158
- [40] C. Wing, K. Tan, L. Yi, The Use of Profits as Opposed to Conventional Forecast Evaluation Criteria to Determine the Quality of Economic Forecasts, Nanyang Business School, Nanyang Technological University, Nanyang Avenue, Singapore, 2007.
- [41] L. Woodland, B. Woodland, Market efficiency and the favourite-longshot bias: the baseball betting market, *The Journal of Finance* 49 (1994) 269–279.