Probability, Risk and Uncertainty in an Inefficient Association Football Gambling Market

ANTHONY C. CONSTANTINOU

E-mail: anthony@constantinou.info - Web: http://www.constantinou.info

Supervision: Prof. Norman Fenton (primary) and Prof. Martin Neil (secondary)

Risk Assessment and Decision Analysis Research Group, Risk Information Management Group, Department of Electronic Engineering and Computer Science, Queen Mary, University of London, UK, E1 4NS

ABSTRACT: Researchers have witnessed the great success in deterministic and perfect information domains. Intelligent pruning and evaluation techniques have been proven to be sufficient in providing outstanding intelligent decision making performance. However, processes that model uncertainty and risk for real-life situations have not met the same success. Association Football has been identified as an ideal and exciting application for that matter; it is the world's most popular sport and constitutes the fastest growing gambling market at international level. As a result, summarising the risk and uncertainty when it comes to football outcomes has been dramatically increased both in importance as well as in challenge.

This study exhibits evidence of an (intended) inefficient football gambling market, presents a rating system for providing appropriate measures of superiority between adversaries, demonstrates a probabilistic Bayesian belief network with forecasting capability that is sufficient for generating profit against published market odds, and deals with propriety in the assessment of probabilistic football match forecasts.

FOOTBALL GAMBLING MARKET [4]

- A gambling market is described as being inefficient if there are one or more betting strategies that generate profit, at a consistent rate, as a consequence of exploiting market flaws
- accuracy between bookmakers is extremely consistent
- bookmakers' accuracy has not improved over the last decade
- profit margins have been dramatically reduced over the last decade
- there are regular predetermined biases in published odds
- there are arbitrage opportunities
- profit margins can be significant even when considering one bookmaker and one football division
- published odds of one bookmaker cannot be considered as representative of the overall market
- systematic adjustments of published odds cannot be explained by rational qualitative factors
- conflicting adjustments of published odds occur between bookmakers
- this inefficiency is considered as the outcome of commercial objectives rather than lack of ability
- the gambling market appears to allow exposure and losses against the very best of bettors and in return increases profits against the residual, more casual bettors

PI-FOOTBALL [2, 3]

- Bayesian network model for forecasting Association Football matches
- subjective variables represent the factors that are important for prediction but which historical data fails to capture
- generates forecasts about the outcomes of the English Premier League matches
- forecasts are published online at www.pi-football.com prior to the start of each match
- for season 2010/11 and at standard discrepancy levels of ≥5% the profitability ranges from **2.87%** to **9.48%**, depending on various bookmakers' odds
- at higher discrepancy levels (8% to 11%) the maximum profit observed ranges from **8.86%** to **35.63%**
- no other published work appears to be particularly successful at beating all of the various bookmakers' odds over a long period of time
- the model continues to be profitable at similar rates for current season 2011/12.

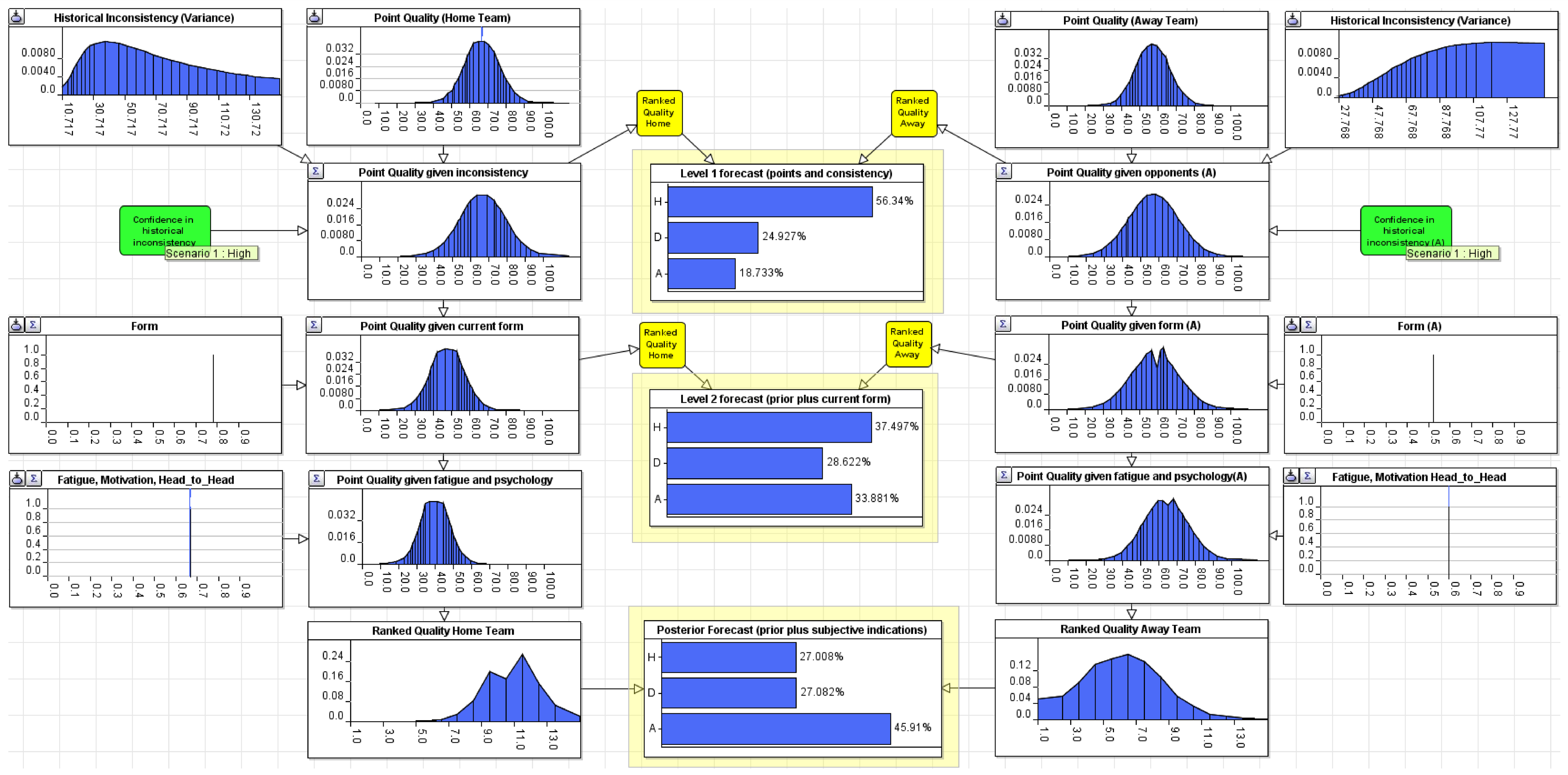

Figure 2.2. Simplified example of the actual Bayesian model used to forecast Arsenal vs. Liverpool, August 20th, 2011. Final result: 0-2

PI-RATING [5]

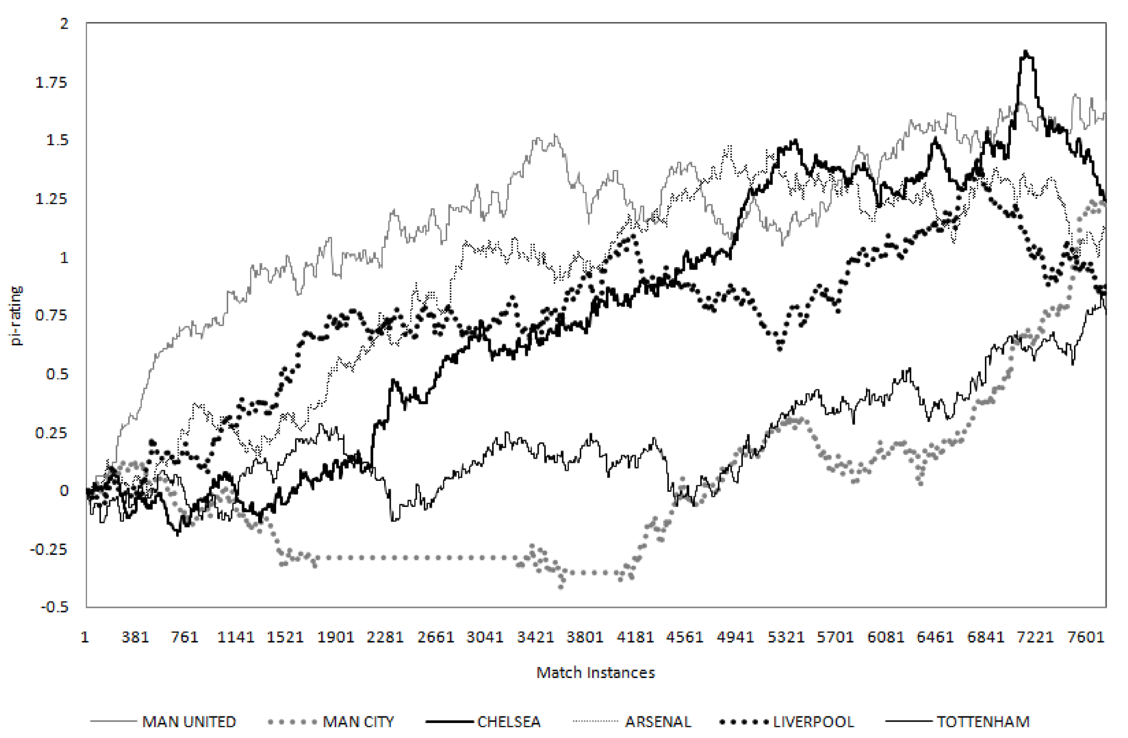

Figure 3.1. Rating development over a period of ~20 seasons for the six most popular EPL teams (from season 1992/93 up to 29th round of season 2011/12).

- provides relative measures of superiority between adversaries
- measured dynamically
- based on the relative discrepancies in scores between adversaries
- ratings are meaningful in terms of expected score difference
- applicable to any other sport (assuming scores)
- outperforms the standard bivariate Poisson distribution method that has been previously proposed and extensively used throughout the football academic literature
- resulting ratings can be used as one of the model parameters for prediction purposes
- allows extensions of this research to answer questions such as: a) *'which football league is best; the English Premier League or the Spanish La Liga?'*, b) *'to what degree lower divisions differ from higher divisions in England'*, c) *how much damage has the 2006 Italian football scandal caused to Serie A?'*.

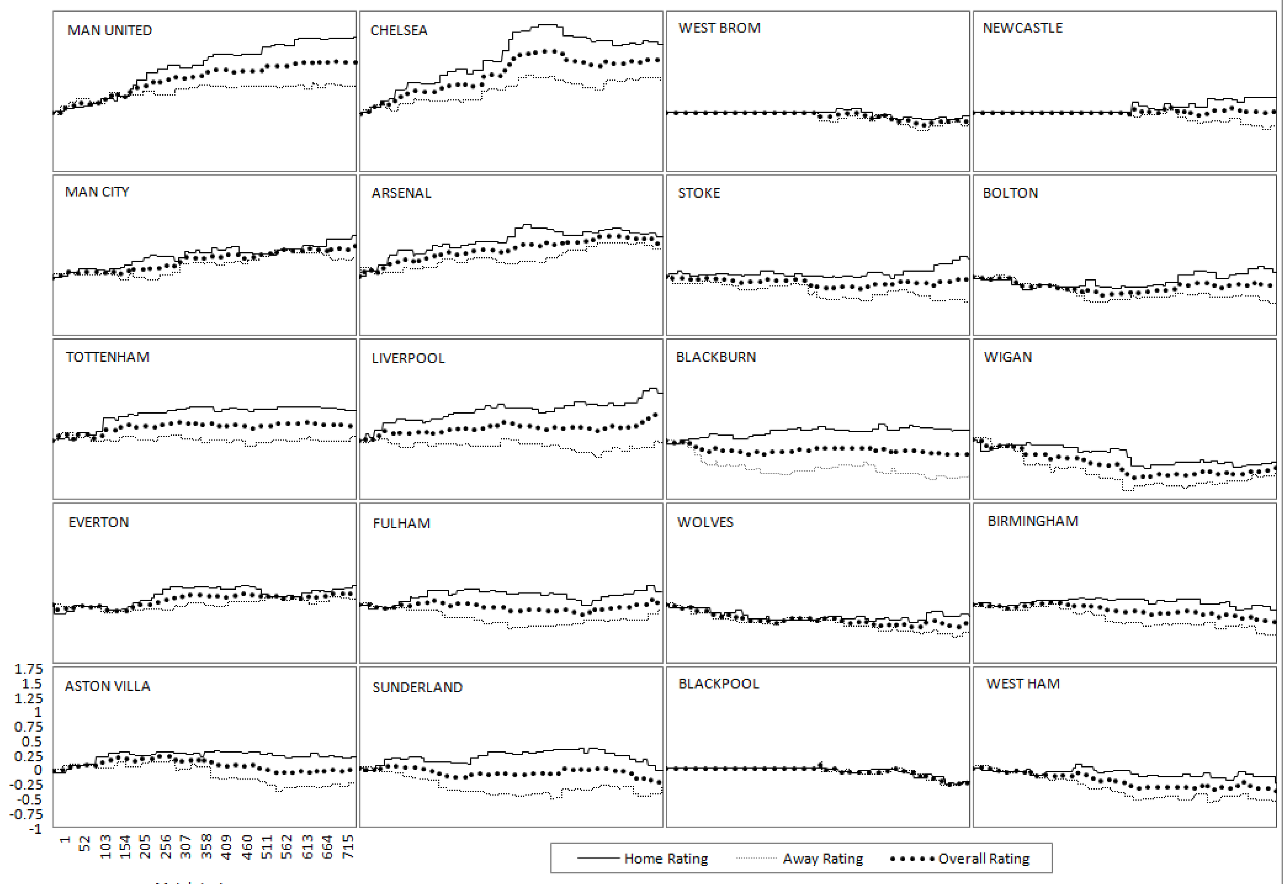

Figure 3.2. Development of the pi-ratings based on individual home and away performances for the specified teams, for seasons 2009/10 and 2010/11.

ACCURACY ASSESSMENT [1]

Match	Model	p(H)	p(D)	p(A)	Result	Best model
1	α	1	0	0	H	α
	β	0.9	0.10	0		
2	α	0.8	0.10	0.10	H	α
	β	0.50	0.25	0.25		
3	α	0.35	0.30	0.35	D	α
	β	0.60	0.30	0.10		
4	α	0.60	0.25	0.15	H	α
	β	0.60	0.15	0.25		
5	α	0.57	0.33	0.10	H	α
	β	0.60	0.20	0.20		

Table 1. Hypothetical forecasts

- a *scoring rule* is a method that is used to assess the accuracy of forecasts or the performance of repeated decisions under uncertainty
- previous studies have been using various scoring rules to assess football models
- no agreed scoring rule existed for determining their forecast accuracy
- the various scoring rules used for validation were inadequate since they failed to recognise that football outcomes represent an ordinal scale distribution
- the Rank Probability Score (RPS) [6] represents the difference between the cumulative distributions of forecasts and observations and overcomes this problem

Match (Model)	Binary Decision Score	Brier Score	Geometric Mean Score	Information Loss Score	MLLE Score
1	✓	✓	✓	✓	✓
(α)	1	0	1	0	0
(β)	0	0.0200	0.9000	0.1520	-0.1054
2	✗	✓	✓	✓	✓
(α)	1	0.0600	0.80	0.3219	-0.2231
(β)	1	0.3750	0.50	1	-0.6931
3	✗	✓	✗	✗	✗
(α)	0	0.7350	0.30	1.7369	-1.2039
(β)	0	0.8600	0.30	1.7369	-1.2039
4	✗	✗	✗	✗	✗
(α)	1	0.2450	0.60	0.7369	-0.5108
(β)	1	0.2450	0.60	0.7369	-0.5108
5	✗	✗ ✗	✗ ✗	✗ ✗	✗ ✗
(α)	1	0.3038	0.57	0.8110	-0.5621
(β)	1	0.0240	0.60	0.7369	-0.5108

Table 2. Applying the scoring rules to each benchmark scenario

Match	Model	$\sum_{j=1}^{i=1,2} p_j$	$\sum_{j=1}^{i=1,2} e_j$	RPS
1	α	1,1	1, 1	(0.0000)
	β	0.90,1	1, 1	0.0050
2	α	0.80, 0.90	1, 1	(0.0250)
	β	0.50, 0.75	1, 1	0.1562
3	α	0.35, 0.65	0, 1	(0.1225)
	β	0.60, 0.90	0, 1	0.1850
4	α	0.60, 0.85	1, 1	(0.09125)
	β	0.60, 0.75	1, 1	0.11125
5	α	0.57,0.90	1, 1	(0.09745)
	β	0.60,0.80	1, 1	0.1

Table 3. RPS scores for each hypothetical forecast

Suppose that for a particular match we have two forecasts f1 and f2:

a) f1 predicts {0.80, 0.15, 0.05} – this represents a strong belief the home team will win and intuitively suggests a comfortable victory,
b) f2 predicts {0.50, 0.30, 0.20} – this represents a weak belief the home team will win and intuitively suggests a narrow victory.

If we know only that the home team won, then any scoring rule is clearly expected to identify forecast f1 as more accurate than f2. If, additionally, we know the home team won 5-0, then again this makes us certain that forecast f1 is superior to f2. However, if the home team won 3-2 then perhaps by intuition this should lead us to conclude that f2 might be the superior forecast.

- Forecasts generated by football models are most typically dependent on historical outcomes of match instances in terms of either a) win-draw-lose and/or b) the number of goals scored by each team
- under either scenario, forecasts are eventually stated and assessed for accuracy in the simplified distribution form of (a)
- in the case of (b), such a forecast distribution is formed by aggregating the prior distribution of goals scored and conceded over all outcomes
- considering only type (a) outcomes to assess football models which were originally generating type (b) forecasts (and hence ignoring the goal difference) inevitably leads to a less accurate forecast assessment
- there is a significant difference between the accuracy scores of forecast types (a) and (b) (for identical match instances)
- ignoring type (b) outcomes for assessment may lead to invalid conclusions about the forecast accuracy of probabilistic football match forecasts.

REFERENCES

[1] Constantinou, Anthony C. and Fenton, Norman E. (2012). Solving the Problem of Inadequate Scoring Rules for Assessing Probabilistic Football Forecast Models. *Journal of Quantitative Analysis in Sports*: Vol. 8: Iss. 1, Article 1.

[2] Constantinou, Anthony C., Fenton, Norman E. and Neil, Martin (2012). pi-football: A Bayesian network model for forecasting Association Football match outcomes. *Knowledge-Based Systems*, 36: 322-339.

[3] Constantinou, Anthony C., Fenton, Norman E. and Neil, Martin (2013). Profiting from an Inefficient Association Football Gambling Market: Prediction, Risk and Uncertainty using Bayesian Networks. *Knowledge-Based Systems*, 50: 60-86.

[4] Constantinou, Anthony C. and Fenton, Norman E. (2013). Profiting from arbitrage and odds biases of the European gambling market. *The Journal of Gambling Business and Economics*, 7 (2): 41-70.

[5] Constantinou, Anthony C. and Fenton, Norman E. (2013). 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*, 9 (1): 37-50.

[6] Epstein, E. (1969). A Scoring System for Probability Forecasts of Ranked Categories. *Journal of Applied Meteorology*, 8, 985-987.

Research Open Day, Queen Mary, University of London, April 17, 2012