Bayesian networks for prediction, risk assessment and decision making in an inefficient Association Football gambling market

By:

Anthony Costa Constantinou

Submitted for the degree of Doctor of Philosophy, 2012

Risk & Information Management (RIM) Research Group,
Faculty of Science and Engineering,
Department of Electronic Engineering and Computer Science,
Queen Mary, University of London,
Mile End Road, London, UK, E1 4NS
E-mail: anthony@constantinou.info
Web: www.constantinou.info

Declaration

I certify that this thesis, and the research to which it refers, are the product of my own work, and that any ideas or quotations from the work of other people, published or otherwise, are fully acknowledged in accordance with the standard referencing practices of the discipline.

______________________________	31/05/2012
Anthony Costa Constantinou	Date

Acknowledgement

I would like to take this opportunity and express my sincere gratitude to my supervisor Professor Norman Fenton, Head of the Risk & Information Management (RIM) Research Group. He is a brilliant supervisor, and when words are not enough I tend to use numbers; I rate his overall support with 10 out of 10.

Special thanks to Professor Martin Neil whose bright hints and expertise have always been very helpful.

Further, I would like to acknowledge the financial support by the Engineering and Physical Sciences Research Council (EPSRC) for fully funding this research, and Agena Ltd for software support.

EPSRC
Engineering and Physical Sciences
Research Council

agena
Bayesian Network
and Simulation Software
for Risk Analysis and
Decision Support

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 the outcomes of relevant football match events has been dramatically increased both in importance as well as in challenge.

A gambling market is described as being inefficient if there are one or more betting procedures that generate profit, at a consistent rate, as a consequence of exploiting market flaws. This study exhibits evidence of an (intended) inefficient football gambling market and demonstrates how a Bayesian network model can be employed against market odds for the gambler's benefit. A Bayesian network is a graphical probabilistic model that represents the conditional dependencies among uncertain variables which can be both objective and subjective. We have proposed such a model, which we call pi-football, and used it to generate forecasts for the English Premier League matches during seasons 2010/11 and 2011/12. The proposed subjective variables represent the factors that are important for prediction but which historical data fails to capture, and forecasts were published online at www.pi-football.com prior to the start of each match.

For assessing the performance of our model we have considered both profitability and accuracy measures and demonstrate that subjective information improved the forecasting capability of our model significantly. Resulting match forecasts are sufficiently more accurate relative to market odds and thus, the model demonstrates profitable returns at a consistent rate.

To my wife Eleni

Glossary of Abbreviations

BN	Bayesian Network
CPD	Conditional Probability Distribution
DAG	Directed Acyclic Graph
EPL	English Premier League
Football	Association Football
NPD	Node Probability Table
OO	Object Oriented
OOBN	Object Oriented Bayesian Network
OOP	Object Oriented Programming
RPS	Rank Probability Score

Contents

Tables

Figures

CHAPTER 1

An introduction to this research project and its objectives

This chapter provides a motivating introduction about the thesis, along with its research hypothesis, its structure, and information on publications, or paper submitted for publication, as a result of this research project.

1.1 Introduction

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. Fenton and Neil demonstrate in

(Fenton & Neil, 2012) how real risks are often misjudged or not taken into consideration by experts in risk assessment. They describe such a real scenario in their book where a show involving tigers was the reason for bringing down the *Mirage Hotel/Casino* on October 3rd 2003, resulting in a loss of hundreds of millions of dollars and making it the single worst loss in Las Vegas history. This example shows that even in the case of casinos where millions are spent on risk assessment and risk management for protecting themselves against any predictable risks, experts have still failed to predict (or even consider) for assessment such a thread which resulted in the largest possible loss they could have ever suffered.

Association Football has been identified as an ideal and exciting application for evaluating probabilistic modelling techniques; its enormous popularity which constitutes it as the most popular sport at international level (Dunning & Joseph A. M., 1993; Mueller et al., 1996; Dunning E., 1999), along with increasing interest in gambling (particularly after its introduction online) means that great attention is now paid to football betting odds. As a result, summarising the risk and uncertainty when it comes to the outcomes of football match events has been dramatically increased both in importance as well as in challenge. While betting interest in horse racing has decreased, betting on football has increased so that is now by far the biggest sport in terms of turnover through the online bookmakers (Finnigan & Nordsted, 2010). *Global Betting and Gaming Consultants* (2001) reported a turnover close to $£2bn$ for British football bookmakers in 1998, and football betting was described as the fastest growing sector in British gambling in (Mintel Intelligence Report, 2001). Unsurprisingly, the turnover reported by just a single bookmaker (bwin Group, 2009) in 2008 was approximately $£2.92bn$; which in turn represented an astonishing 31.4% increase from the year before.

For any large scale gambling market (and this includes financial markets) the question of efficiency is paramount. If there is a betting procedure that is consistent in generating profit against a gambling market, then such a market is normally described as being inefficient. Indeed, the possibility from profiting because of market flaws is usually what makes such studies both important and exciting. Because of the explosion of interest in football betting, an increasing number of researchers have turned their attention to evaluating the efficiency of this particular betting market and developing various football forecast models. However, even though numerous evidence of inefficiency has been claimed by many researchers (see Chapter 5), a particularly successful football model that generates profit against the various inefficient bookmakers' odds at a consistent rate is still missing from the published academic literature (see Chapter 3). The vast majority of the previous studies concerned with football match prediction were focused on purely statistical approaches and generated predictions solely on the basis of relevant objective information; implying that important information for prediction (i.e. team motivation and player injuries) that is not captured by the historical data is completely ignored, and no successful attempts appear to have been made to properly incorporate subjective information along with relevant historical data.

In (Joseph et al., 2006) the authors demonstrated how an expert constructed Bayesian network (BN) provided superior performance against various machine learning techniques in predicting the outcome of football matches involving Tottenham Hotspur. BNs are a powerful tool for modelling causality (rather than correlation as standard statistical approaches do) between both objective and subjective variables of interest for prediction, risk assessment and decision making purposes under uncertainty. BNs have already been employed to model knowledge with success in many different fields such as bioinformatics, engineering, law, gaming, medicine and image processing. A novel BN football model that will consider both objective and subjective information for prediction, whereby subjective information will represent information that is important for prediction but which historical data fails to capture, should be able to provide superior forecasting capability, compared to the previously proposed approaches, in an attempt to beat the market.

In (Dixon & Coles, 1997) the authors claimed 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, and (Graham & Stott, 2008) suggested that if such a work was particularly successful, it would not have been published.

This study investigates the efficiency of the Association Football (hereafter referred to simply as 'football') gambling market and demonstrates how Bayesian networks (BNs) can be used to exploit inaccurate published odds in an attempt to generate positive expected returns for the bettor. A BN is a graphical probabilistic model that represents the conditional dependencies among uncertain variables which can be both objective and subjective.

Our proposed BN model, which we call pi-football, generates football match forecasts based on both objective and subjective information, whereby proposed subjective variables represent the factors that are important for prediction but which historical data fails to capture. This study represents the first comprehensive approach to this kind of football predictions, and because of the nature of the subjective information we have been publishing our forecasts online at www.pi-football.com prior to the start of each match; earlier studies which incorporated subjective information have not done so (Joseph et. al., 2006; Min et al., 2008; Baio & Blangiardo, 2010).

In assessing the forecasting capability of our model, we have investigated the previously proposed accuracy measures under relevant literature review studies. In doing so, we have discovered that all of the measures proposed and used are both inadequate and inconsistent in assessing the accuracy of football match forecasts between and odds offered by the various bookmaking firms. In fact, serious concerns over the various forecast accuracy measures have already been exposed in macroeconomics (Leitch & Tanner, 1991; Armstrong & Collopy, 1992; Hendry, 1997; Fildes & Stekler, 2002), and some suggested the use of more than one measure in an attempt to obtain an informed picture of the relative merits of the forecasts (Jolliffe & Stephenson, 2003). The relevance of macroeconomics to gambling markets is that in both cases accuracy and profit are important when assessing economic forecast methods. We have therefore proposed a well-established forecast measure for that matter, but we have considered both accuracy and profitability measures for assessing the performance of our model since earlier studies from macroeconomic domains have shown conflicting conclusions between the two (Leitch & Tanner, 1991), whereas others have concluded that it might be best to combine profitability methodologies with a proper forecast assessment method (Wing et al., 2007).

We have assessed the forecasting capability of our model over two continuous English Premier League (EPL) seasons (2010/11 and 2011/12), where the model considered during the latter season was an improved version of the former in an attempt to increase the predictive power as well as to reduce model complexity. Our results demonstrate that subjective information increased the forecasting capability of our model significantly under both seasons, and the model was able to generate positive returns at a consistent rate against all of the (available) bookmakers' odds; a predictive performance that is superior to any other relevant academic published work in football match prediction and betting which highlights the success of pi-football.

1.2 Research hypothesis

The hypothesis of this thesis is that sports gambling markets, and particularly football, publish odds that are biased towards maximising profitability and hence, market odds suffer from a degree of inaccuracy. This intended inefficiency can be exploited with sophisticated probabilistic models that properly incorporate subjective information along with relevant historical data and hence, become sufficiently accurate in an attempt to outperform the market for profit.

All of the previous relevant academic research studies have failed to demonstrate profitability that is consistent over time against published market odds, and the vast majority of these studies are solely focused on data-driven approaches to prediction, and by relying on purely statistical approaches.

In an attempt to provide superior forecasting capability over the previously proposed football models, and to demonstrate profitability against market odds, we propose the development of a novel BN football model to model causality (rather than correlation) between objective and subjective variables of interest; whereby subjective inputs will represent information that is important for prediction but which historical data fails to capture (i.e. fatigue, form, motivation, player availability, influence of a new manager, new player transfers).

1.3 Structure of the thesis

The thesis is organised as follows: Chapter 2 provides background information on Bayesian reasoning, Chapter 3 summarises the current *state-of-the-art* from the published academic literature, Chapter 4 deals with propriety in assessing the forecast accuracy of football forecast models, Chapter 5 provides evidence of an inefficient football gambling market, Chapter 6 demonstrates the initial Bayesian network model used to generate football match forecasts during the EPL season 2010/11 along with the results, Chapter 7 demonstrates the extended Bayesian network model (based on that of Chapter 6) used for to generate football match forecasts during the EPL season 2011/12 along with the results, Chapter 8 presents a system for determining the level of ability of football teams by dynamic ratings based on the relative discrepancies in scores between adversaries whereby resulting ratings can be used as an input to the Bayesian network model in an attempt to further enhance its forecasting capability, and finally we provide a summary of our results, along with potential future directions, in Chapter 9.

1.4 Publications and papers submitted for publication

The work in this thesis has led to the following articles:

[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, DOI: 10.1515/1559-0410.1418 (Draft available online at: http://constantinou.info/downloads/papers/solvingtheproblem.pdf)

[2] Constantinou, Anthony C. and Fenton, Norman E. (2012). Evidence of an (intended) inefficient Association Football gambling market. Submitted for publication. (Draft available online at: http://constantinou.info/downloads/papers/evidenceofinefficiency.pdf)

[3] Constantinou, Anthony C., Fenton, Norman E. and Neil, Martin. (2012). pi-football: A Bayesian network model for forecasting Association Football match outcomes. To appear in Knowledge-Based Systems, 2012. Online publication August 21, 2012, DOI: 10.1016/j.knosys.2012.07.008 (Draft available online at: http://constantinou.info/downloads/papers/pi-model11.pdf)

[4] Constantinou, Anthony C., Fenton, Norman E. and Neil, Martin. (2012). Profiting for an inefficient Association Football gambling

market: Prediction, Risk and Uncertainty using Bayesian networks. Submitted for publication, 2012. (Draft available online at: http://constantinou.info/downloads/papers/pi-model12.pdf)

[5] Constantinou, Anthony C. and Fenton, Norman E. (2012). Determining the level of ability of football teams by dynamic ratings based on the relative discrepancies in sores between adversaries. Submitted for publication. (Draft available online at: http://www.constantinou.info/downloads/papers/pi-ratings.pdf)

CHAPTER 2

An Introduction to Bayesian Reasoning

This chapter provides an introduction to the Bayes' theorem; the theorem whereby this research is based on for prediction, risk management and decision making purposes. We then demonstrate how BNs take advantage of this theorem for modelling uncertain variables and introducing causal relationships between them.

2.1 Bayes' theorem

Bayes' theorem is just a simple equation which relates conditional and marginal probability distributions of random variables, and this theorem was named after the English Mathematician Reverend Thomas Bayes who, in the late 1750s, studied how to compute a distribution for the probability parameter of a binomial distribution and how prior beliefs can be updated based on new evidence; a process that we now call Bayesian inference whereby belies are stated as *prior* probabilities and updated beliefs are stated as *posterior* probabilities. After Bayes' death in 1761, his friend Richard Price edited and published Bayes' work in 1763 as "An Essay towards solving a Problem in the Doctrine of Chances" (Bayes, 1763), which served as the first detailed description based on probability theory.

The Bayes' equation below shows that however different the probability of event A conditional upon event B is to that of B conditional upon A, there is still a relationship between the two.

$$p(A|B) = \frac{p(B|A) \times p(A)}{p(B)}$$

In many cases the event space $\{A_i\}$ is specified for finite partitions of the event space in terms of $p(A_i)$ and $p(B|A_i)$. Under such cases it is useful to eliminate $p(B)$ using the *law of total probability* and define an extended form of the Bayes' theorem such that:

$$p(B) = \sum_i p(B|A_i)p(A_i) \quad \Rightarrow \quad p(A_i|B) = \frac{p(B|A_i)p(A_i)}{\sum_i p(B|A_i)p(A_i)}$$

2.1.1 Paradigm 1: The Monty Hall problem

The famous Monty Hall problem can be solved by using the Bayes' theorem equation. This problem states that there are three doors; a blue, a yellow and a red door. One of these doors has a prize hidden behind it whereas the other two doors have nothing. Suppose we choose the blue door. The presenter, who knows where the prize is, will then open one of the remaining two doors. The presenter will always open a door that has no prize behind it. For this situation, he opens the red door revealing that there is no prize behind it and he asks if we wish to change our initial selection of the blue door. At this point, the problem we are facing is to whether a change of our initial selection will have any impact on our chances of winning the prize. Should the contestant switch doors? Does it really matter? To solve this problem we can use the Bayes' theorem. Accordingly, we define the following variables:

- Let A_β = the event that the prize is behind the blue door.
- Let A_γ = the event that the prize is behind the yellow door.
- Let A_λ = the event that the prize is behind the red door.

$$\text{Thus, } p(A_\beta) = p(A_\gamma) = p(A_\lambda) = \frac{1}{3}$$

- Let B= "*the presenter opens the red door*".

Scenario 1: The prize is behind the blue door. The presenter is free to pick between the yellow and the red door.

$$\text{Thus, } p(B|A_\beta) = \frac{1}{2}$$

Scenario 2: The prize is behind the yellow door. The presenter must pick the red door.

$$\text{Thus, } p(B|A_\gamma) = 1$$

Scenario 3: The prize is behind the red door. The presenter must pick the yellow door.

$$\text{Thus, } p(B|A_\lambda) = 0$$

By applying the values to the Bayes' equation we get the following results:

$$p(A_\beta|B) = \frac{p(B|A_\beta) \times p(A_\beta)}{p(B)} = \frac{\frac{1}{2} \times \frac{1}{3}}{\frac{1}{2}} = \frac{1}{3}$$

$$p(A_\gamma|B) = \frac{p(B|A_\gamma) \times p(A_\gamma)}{p(B)} = \frac{1 \times \frac{1}{3}}{\frac{1}{2}} = \frac{2}{3}$$

$$p(A_\lambda|B) = \frac{p(B|A_\lambda) \times p(A_\lambda)}{p(B)} = \frac{0 \times \frac{1}{3}}{\frac{1}{2}} = 0$$

The Bayes' theorem shows that we should always switch, under such a scenario, in order to maximise our chances of winning the prize. In particular, the probability that the prize is behind the red door is $\frac{1}{3}$ whereas the probability that the prize is behind the blue door is $\frac{2}{3}$.

The critical information here is that the presenter knows where the prize is, so he will always open an empty door. Since he is not opening a door at random, the revised probability that the prize is behind the door which you have initially chosen stays unaltered at $\frac{1}{3}$ and thus, the revised probability for the prize to be behind the blue door has to be $\frac{2}{3}$.

2.1.2 Paradigm 2: Medical Doctors and Probabilistic Reasoning (Disease test)

The following problem was put by Casscells, Schoenberger and Grayboys (1978) to 60 students and staff at Harvard Medical School:

"if a test to detect a disease whose prevalence is $\frac{1}{1000}$ *has a false positive rate of* 5%*, what is the chance that a person found to have a positive result actually has the disease, assuming you know nothing about the person's symptoms or signs?"*

Using the Bayes' theorem, we define the following variables:

- Let A = "the event that the person has the disease".
- Let B = "the event that the test is positive".

Thus, $p(A) = 0.001$ and $p(\neg A) = 1 - A = 0.999$,

and also $p(B|A) = 1$ and $p(B|\neg A) = 0.05$

We want to know the value of A; what is the probability that a person has the disease, given the value of B; that the test associated with that person was positive with regards to the disease. Using the Bayes' theorem we get:

$$p(A|B) = \frac{p(B|A) \times p(A)}{p(B)} = \frac{p(B|A) \times p(A)}{p(B|A) \times p(A) + p(B|\neg A) \times p(\neg A)}$$

We assign the known values to each of the variables and get:

$$\frac{p(B|A) \times p(A)}{p(B|A) \times p(A) + p(B|\neg A) \times p(\neg A)} = \frac{1 \times 0.001}{1 \times 0.001 + 0.05 \times 0.999} = 0.0196$$

Assuming that the probability of a positive test result given that the person has the disease is 1, then the answer is approximately 2% as demonstrated above. (Casscells et al., 1978) showed that only 18% of the participants gave this answer, whereas the modal response was 95%; presumably because of the error rate of the test is 5% and therefore it must return 95% correct results.

2.2 Bayesian networks

To begin with it is important to note that BNs are often known by other names and most notably these include: influence diagrams (Shachter, 1986), causal probabilistic networks (Jensen et al., 1990), recursive graphical models (Lauritzen,1995), Bayesian belief networks (Cheng et al., 1997), belief networks (Darwiche, 2002) and causal networks (Heckerman, 1995a; 2007). Even though authors might often mean slightly different things when they use the above or any other similar terms, the term Bayesian network appears to have become the prevalent way of describing this kind of structured modelling (Daly et al., 2011). There are many standard books which cover the theory of BNs and these include (Pearl, 2000; Jensen, 2001; Neapolitan, 2004; Fenton & Neil, 2012), as well as short tutorials for a quick introduction (Heckerman, 1995b).

A BN is a graphical probabilistic model that represents the conditional dependencies among uncertain variables which can be both objective and subjective, whereby random variables are represented by nodes and causal influences are represented by arrows. BNs are powerful tools for decision support systems and have become increasingly recognised as a potentially powerful solution to complex risk assessment problems (Heckerman et al., 1995). BNs have already been employed to model knowledge in many different fields such as computational biology and bioinformatics (Friedman et al., 2000; Jiang et al., 2011), engineering (Pourret & Marcot, 2008), computer science (Fenton & Neil, 2004; Pourret & Marcot, 2008), artificial intelligence and machine learning (de Campos et al., 2004; Koumenides & Shadbolt, 2012), law (Davis, 2003; Kadane & Schum, 1996), gaming, business gambling, natural sciences (Pourret & Marcot, 2008), medicine (Uebersax, 2004; Pourret & Marcot, 2008; Jiang & Cooper, 2010) and image processing (Diez et al., 1997).

2.2.1. Bayesian network software

Several commercial and non-commercial software tools exist for developing BNs. The models demonstrated within this thesis have been developed using AgenaRisk (Agena, 2012), a commercial BN and simulation software for risk analysis and decision support. Agena Ltd provides software support for our research group, *Risk & Information Management* (RIM)[1]. The most important differentiator between AgenaRisk and other BN tools is its ability to properly incorporate continuous variables, without any constraint, and without the need for static discretisation. It does this through its dynamic discretisation algorithm that produces results with far greater accuracy than is possible otherwise. (Neil et al., 2010). Other BN software tools or packages include:

[1] Formerly RADAR: Risk Assessment and Decision Analysis Research Group

- @RISK: a commercial risk analysis software using Monte Carlo simulation for Excel (Palisade, 2012);
- Analytica: a commercial influence diagram-based software, for both Windows and Macintosh, with a visual environment for creating and analysing probabilistic models (Lumina Decision Systems, 2012);
- AT-Sigma Data Chopper: a commercial database analysis software for finding causal relationships (AT Sigma, 2007);
- BAYDA: A non-commercial BN software package which is based on the Naive Bayes classifier and allows for strong independence assumption between features (Kontkanen et. al., 1998);
- Bayesian Network tools in Java (BNJ): a non-commercial open-source suite of Java tools for probabilistic learning and reasoning (Hsu, 2004);
- BayesianLab: a commercial analysis toolbox with a complete set of BN tools that include supervised and unsupervised learning (Bayesia, 2001);
- Bayes Server: a commercial advanced BN software library and user interface supporting classification, regression, segmentation, time series prediction, anomaly detection and more (Bayes Server, 2012);
- Bayesware Discoverer: a commercial automated modelling tools that is able to extract a BN model from data by searching for the most probable model (Bayesian Knowledge Discoverer, 1998);
- BNet: a commercial software that includes BNet.Builder for rapid BN development, input and results; and BNet.EngineKit for incorporating Belief Network Technology in your applications (BNET, 2004);
- DXpress: a commercial Windows-based software for building and compiling BNs (Knowledge Industries, 2006);

- FDEP: a non-commercial software for inducing functional dependencies from relations (Flach & Savnik, 1999; FDEP, 2001);
- Flint: a commercial BN software which incorporates fuzzy logic and certainty factors within a logic programming rules-based environment (Logic Programming Associates, 2012);
- GeNle: a non-commercial versatile and user-friendly development environment for graphical decision theoretic models (Decision Systems Laboratory, 2005);
- HUGIN: a commercial software with a full suite of BN reasoning tools (HUGIN EXPERT, 1989);
- J Cheng's Bayesian Belief Network Software: a non-commercial BN software that includes the *BN PowerConstructor*; an efficient system that learns Bayesian belief network structures and parameters from data, and *BN PowerPredictor*; a data mining system for data modelling, classification and prediction (Cheng, 2001);
- JavaBayes: a non-commercial BN system that allows the user to import, create, modify and export such networks (JavaBayes, 2001);
- jBNC: A non-commercial Java toolkit for training, testing and applying BN classifiers (jBNC Toolkit, 2004);
- JNCC2, Naive Credal Classifier 2: an non-commercial classifier that constitutes an extension of the traditional Naive Bayes Classifier towards imprecise probabilities, and it is design to return robust classification for even small and/or incomplete data sets (Corani & Zoffalon, 2008; JNCC2, 2008);
- MSBNx: Microsoft Belief Network Editor: a component-based Windows application for creating, assessing, and evaluating BNs, createrd at

Microsoft Research, and which is provided free to non-commercial research users (MSBNx, 2012);

- Netica: a commercial BN and influence diagram software (Norsys Software Corp., 1995);
- PNL: Open Source Probabilistic Networks Library: a non-commercial tool for working with graphical models. It supports directed and undirected models, discrete and continuous variables, and various inference and learning algorithms (PNL, 2010);
- PrecisionTree: a commercial add-on for Microsoft Excel which allows the development of decision trees and influence diagrams directly in the spreadsheet (Palisade, 2012);
- Pulcinella: a non-commercial tool, written in CommonLisp, for propagating uncertainty through local computations based on the general frameworks of valuation systems (Pulcinella, 1996) proposed by (Shenoy & Shafer, 1988);
- SMILE (Structural Modeling, Inference, and Learning Engine): a non-commercial fully portable library of C++ classes implementing graphical decision theoretic methods directly amenable to inclusion in intelligent systems (Decision Systems Laboratory, 2005).

2.2.2. A Bayesian network example: Revisiting the Monty Hall problem

In this section we revisit the Monty Hall Problem introduced in Section 2.1.1 and demonstrate how this can be solved using BNs. We provide two different solutions based on those presented in (Fenton & Neil, 2012); one simple and one complex. Both solutions give the same results, but the complex solution better illustrates the causal structure of the problem.

- Simple Solution:

The simple BN model in its initial state is shown in Figure 2.1, and the node probability table (NPT) for the node *Door shown empty* is given in Table 2.1.

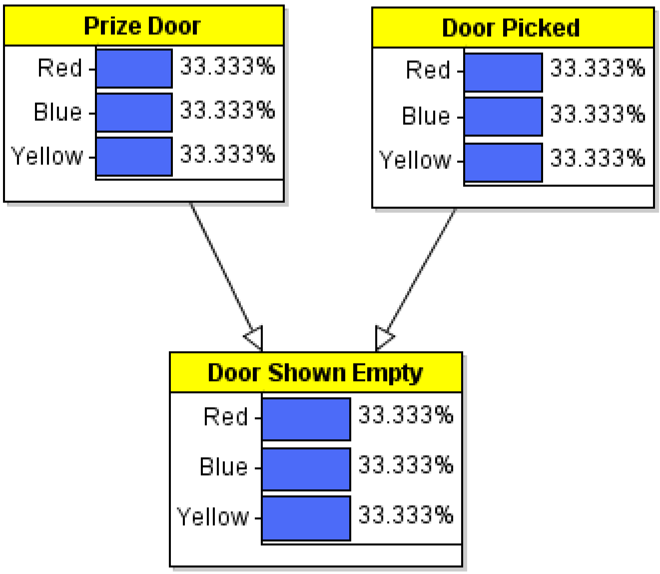


Figure 2.1. Simple solution: Structure of the model and its initial state.

Table 2.1. Simple solution: The NPT for the node *Door shown empty.*

Prize Door	Red			Blue			Yellow		
Door Picked	Red	Blue	Yellow	Red	Blue	Yellow	Red	Blue	Yellow
Red	0.0	0.0	0.0	0.0	0.5	1.0	0.0	1.0	0.5
Blue	0.5	0.0	1.0	0.0	0.0	0.0	1.0	0.0	0.5
Yellow	0.5	1.0	0.0	1.0	0.5	0.0	0.0	0.0	0.0

After we pick the *Blue* door then, as shown in Figure 2.2, at this point each door is still equally likely to win. However, when the presenter opens the red door (which will always be empty) moves all of the probability that was previously associated with the two doors not chosen by the contestant (red and yellow) into the remaining yellow door (see Figure 2.3). Therefore, if the contestant switches doors he/she will increase the probability to find the prize door from $\frac{1}{3}$ to $\frac{2}{3}$.

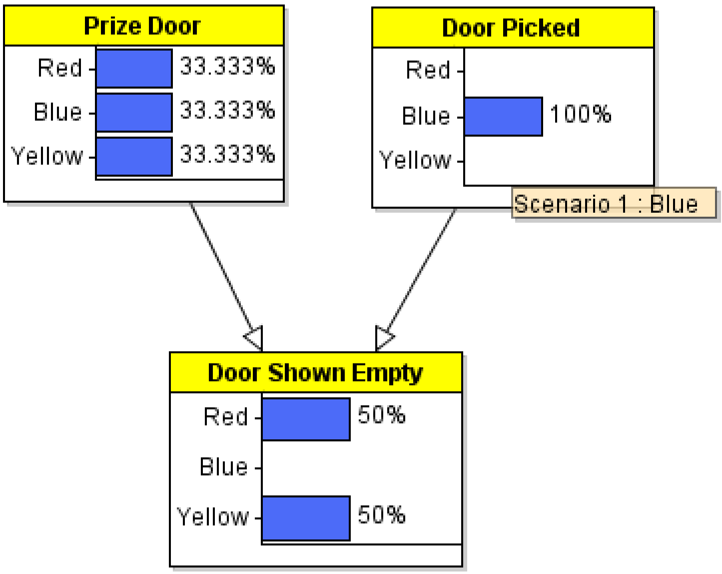


Figure 2.2. Simple solution: The contestant picks the blue door.

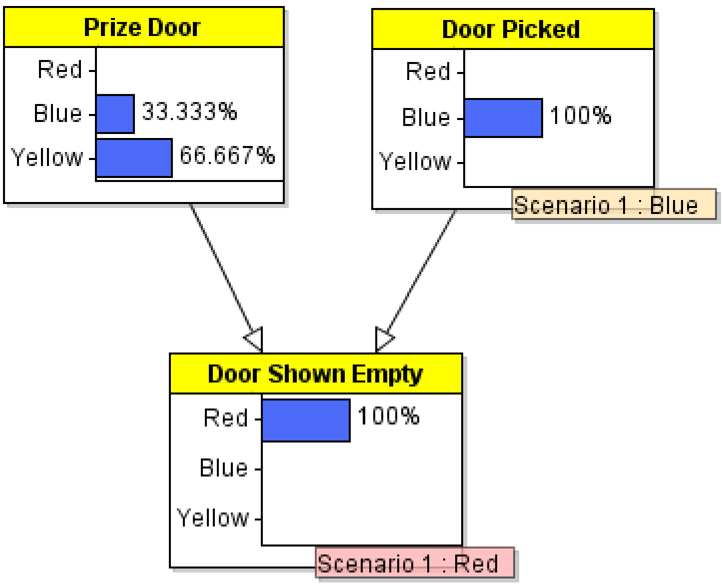


Figure 2.3. Simple solution: The presenter shows the red door empty.

- Complex Solution:

In this solution we demonstrate how the probabilities of the events (nodes) *Switch or Stick*, *Doors after choice*, and *Win Prize* are altered during the game. The initial state of this model is presented in Figure 2.4. The NPT for the node *Door shown empty* is identical to that of the simple model (Table 2.1), whereas the NPT for the node *Door after choice* is shown in Table 2.2, and the NPT for the node *Win Prize* is shown in Table 2.3.

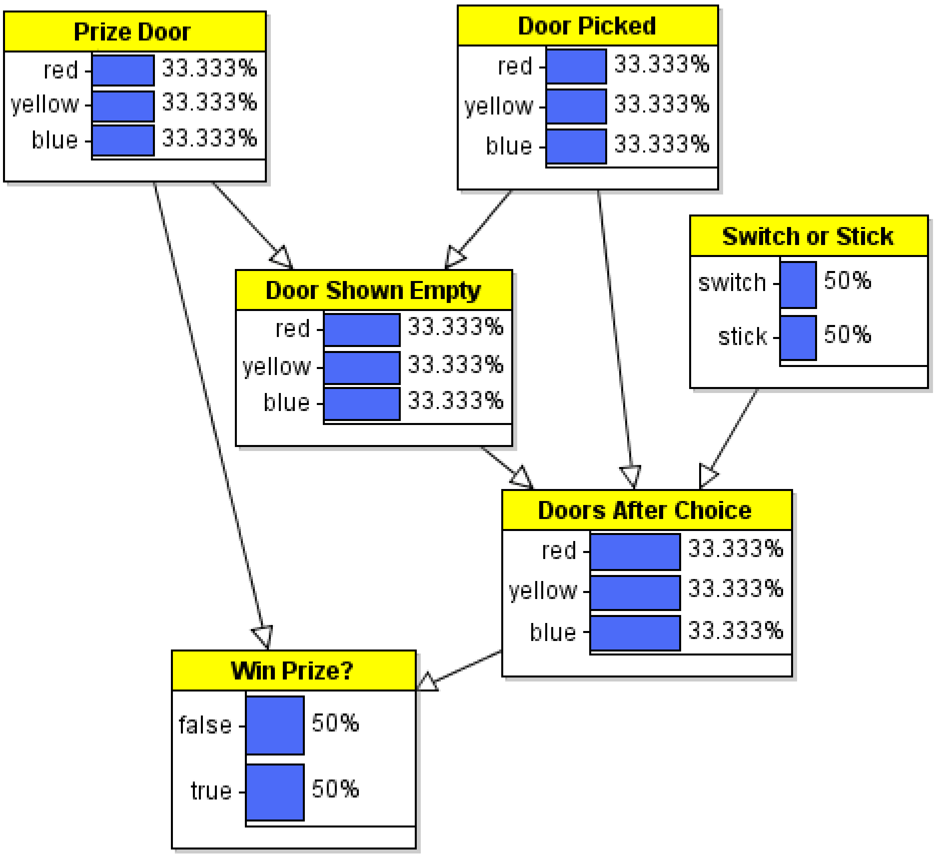


Figure 2.4. Complex solution: Structure of the model and its initial state.

Table 2.2. Complex solution: The NPT for the node *Door after choice*[2].

Door Picked	red						yellow						blue					
Door Show...	red		yellow		blue		red		yellow		blue		red		yellow		blue	
Switch or S...	switch	stick	switch	stick	switch	stick	switch	stick	switch	stick	switch	stick	switch	stick	switch	stick	switch	stick
red	0.3333333	1.0	0.0	1.0	0.0	1.0	0.0	0.0	0.3333333	0.0	1.0	0.0	0.0	0.0	1.0	0.0	0.3333333	0.0
yellow	0.3333333	0.0	0.0	0.0	1.0	0.0	0.0	1.0	0.3333333	1.0	0.0	1.0	1.0	0.0	0.0	0.0	0.3333333	0.0
blue	0.3333333	0.0	1.0	0.0	0.0	0.0	1.0	0.0	0.3333333	0.0	0.0	0.0	0.0	1.0	0.0	1.0	0.3333333	1.0

Table 2.3. Complex solution: The NPT for the node *Win Prize*.

Doors After...	red			yellow			blue		
Prize Door	red	yellow	blue	red	yellow	blue	red	yellow	blue
false	0.0	1.0	1.0	1.0	0.0	1.0	1.0	1.0	0.0
true	1.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0	1.0

As before, suppose that the contestant picks the blue door. Figure 2.5 demonstrates the state of the model after this choice. Note that, since we assume that the choices of *stick* and *switch* are equally likely, the model infers that there is a 50% chance of winning the prize. When the presenter reveals the red door (Figure 2.6) the only information that changes is that the remaining door (yellow) has $\frac{2}{3}$ chance of being the winning door. If the contestant decides to switch doors, it should become clear now that the probability of winning the prize becomes $\frac{2}{3}$ as shown in Figure 2.7.

[2] Note that some of the probabilities assigned to the table correspond to columns which represent impossible events, such as the combination *Door Picked=red* and *Door shown empty=red*. The fact that some combinations are impossible is already encoded in node *Door shown empty* (Table 2.1). However, we still have to assign some probabilistic values to these columns in the NPT, and the standard approach (as adopted in Table 2.2) is to assign equal probability to each column entry (hence the assignment of $\frac{1}{3}$s in Table 2.2) (Fenton & Neil, 2012).

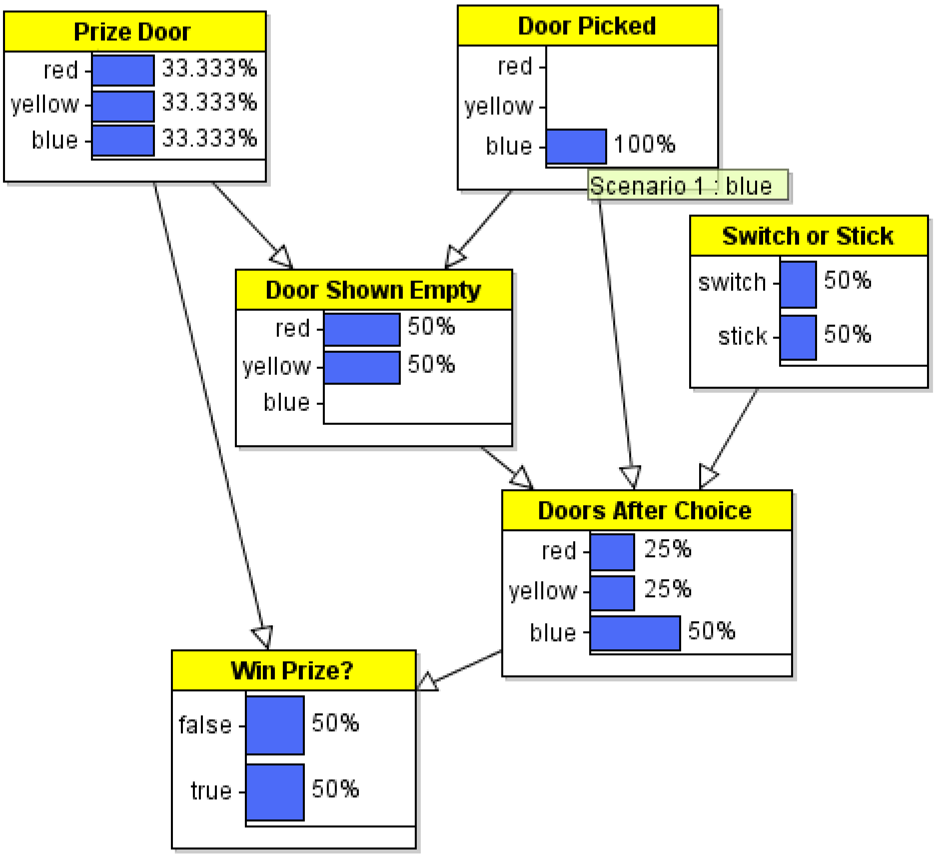


Figure 2.5. Complex solution: The contestant picks the blue door.

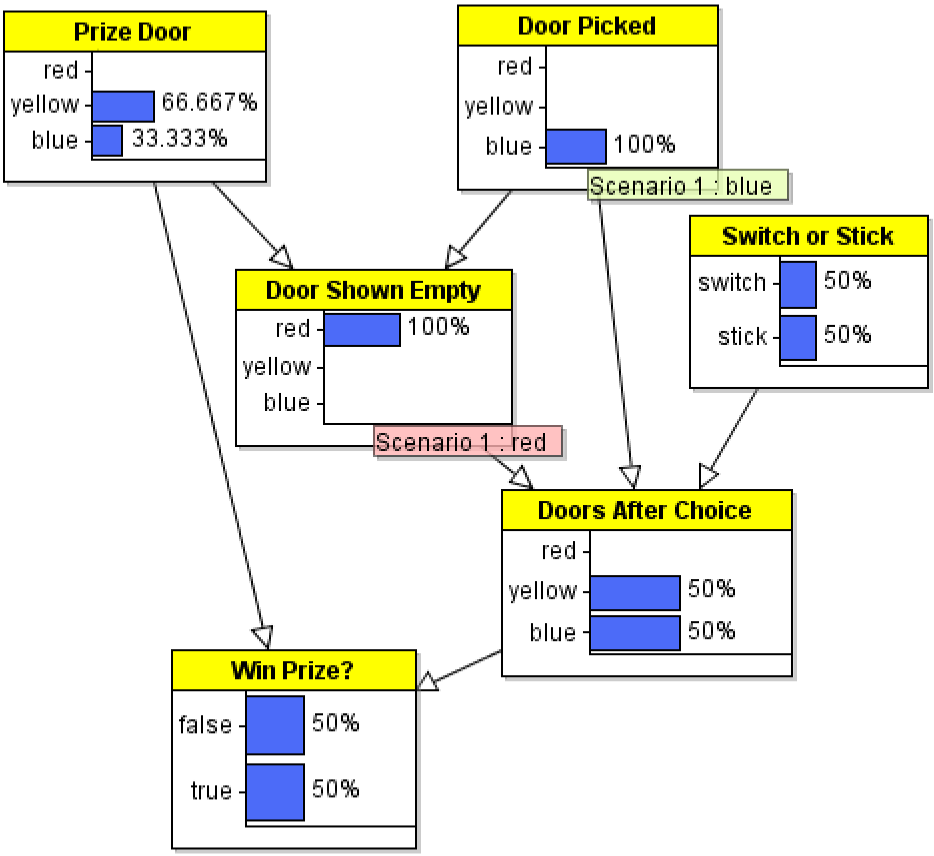


Figure 2.6. Complex solution: The presenter shows the red door empty.

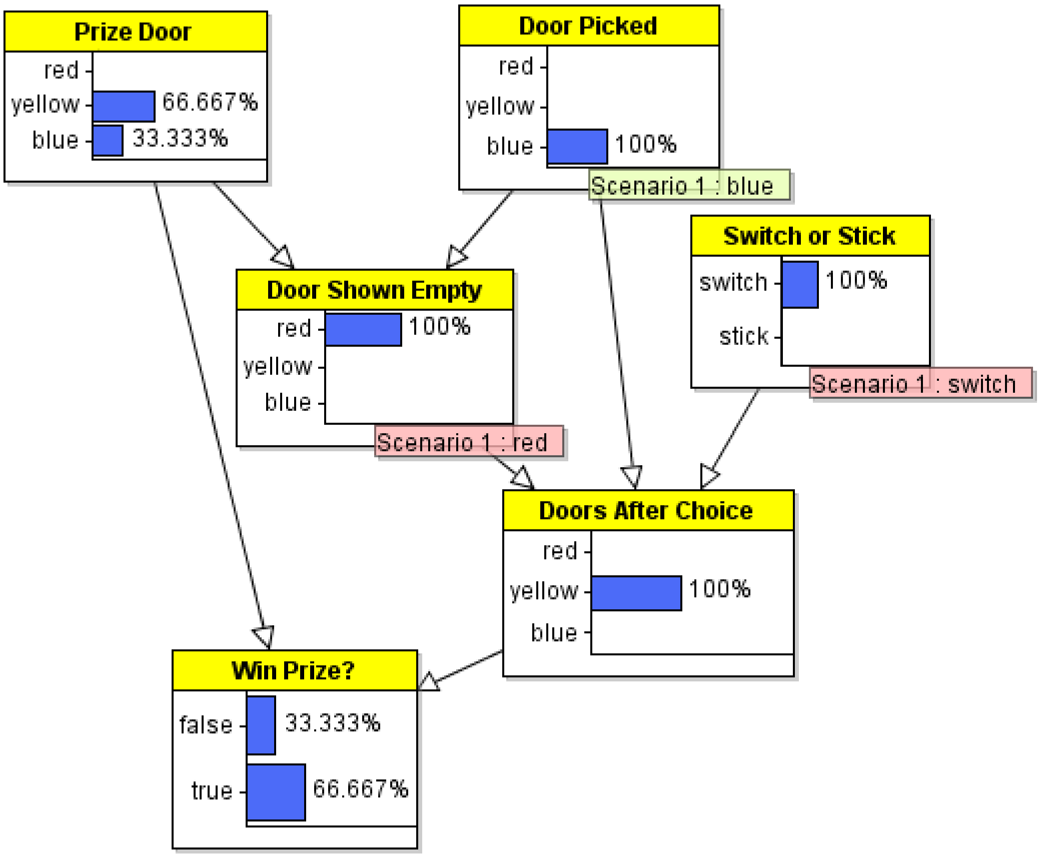


Figure 2.7. Complex solution: The contestant chooses to switch.

2.2.3. Risk Assessment with Bayesian networks and Subjective Information

The aim of this section is to demonstrate the power and flexibility of BNs in handling risk assessment problems, by incorporating subjective information along with relevant historical data, relative to the standard statistical techniques. We use the example of (Fenton & Neil, 2012) based on an automobile crash information as provided by (US Department of Transportation, 2008) in order to explain the need for a BN structure.

Table 2.4 gives the average temperature along with the number of automobile crashes resulting in fatalities in the USA in 2008 as specified by month (US Department of Transportation, 2008), and Figure 2.8 presents the scatterplot graph of temperature against road fatalities. From a quick view there seem to be more fatalities as the temperature increases, so we may conclude that there is a relationship between temperature and fatalities. In fact, according to the *Pearson Correlation Coefficient*[3] formula which is widely used by statisticians, that measures the extent to which the two sets of numbers are related, the correlation coefficient is approximately 0.869. This coefficient considers the chance of a dependency between the datasets to be 'highly significant', comfortably passing the criteria for a p-value[4] of 0.01.

Table 2.4. Temperature and fatal automobile crashes.

Month	Average Temperature	Total Fatal Crashes
January	17	297
February	18	280
March	29	267
April	43	350
May	55	328
June	65	386
July	70	419
August	68	410
September	59	331
October	48	356
November	37	326
December	22	311

[3] The *correlation coefficient* is a number between -1 and $+1$ that determines whether two paired datasets are related. It measures the strength of linear dependence such that we are more confident of a positive linear correlation when the value is closer to $+1$ and vice versa. When the correlation coefficient is close to 0 then it gives evidence of no relationship between the two datasets.

[4] The p-value is the standard method that statisticians use to measure the significance of their empirical analysis. The p-value lies between 0 and 1 inclusive and represents the probability that the data would have arisen if the null hypothesis were true.

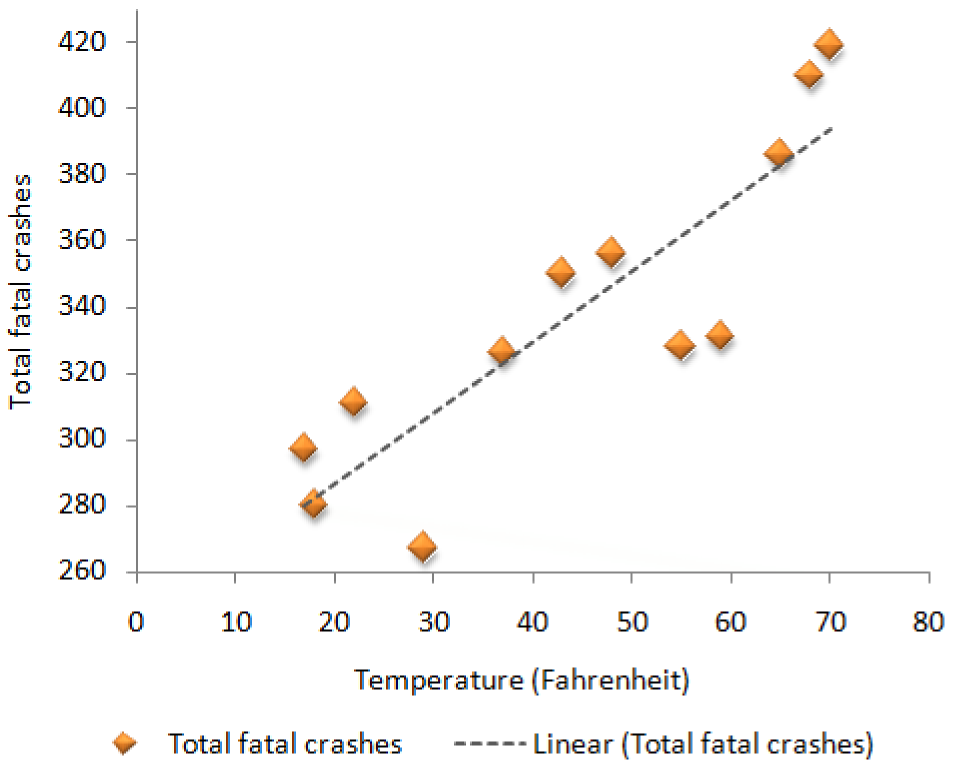


Figure 2.8. Scatterplot of temperature against road fatalities (each dot represents a month).

Based on the above information statisticians might conclude that the number of road fatalities is significantly related to the temperature of any given day. More worryingly, people might infer causal links between the two datasets; for example the higher temperature causes more fatalities. Applying naive statistical regression techniques to this data we will end up with a simple model that looks like the model presented in Figure 2.9, where inference can be measured based on the linear fit of the scatterplot graph in Figure 2.8. This approach will allow us to predict the number of fatal car crashes based on the temperature, and this kind of analysis can lead to dangerous (and counterintuitive) headlines such as *"Driving in Winter is safer than any other time of the year"*.

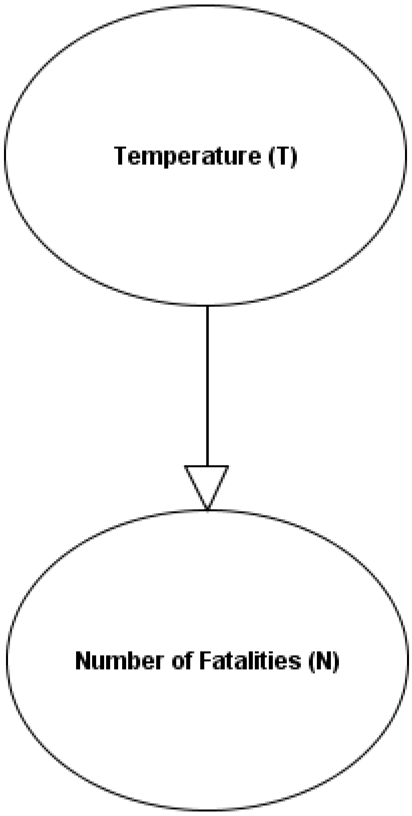


Figure 2.9. Simple regression model for automobile fatalities.

The problem with this example is that there are other, and probably many, underlying factors that contribute to an explanation of the number of road fatalities on any given day and thus, for risk assessment and risk management the regression model is useless since it provides no explanatory power. That is, if we were to perform risk management based on the above example then we would have suggested to do your driving when the fatal car crashes are lower, in Winter. But this is counterintuitive since in Winter the highways are at their most dangerous, and as such common sense suggests that we should expect the risk to increase.

According to (Fenton & Neil, 2012), the causal factors that might do much to explain the apparently strange statistical observations in order to provide better insights into risk are:

- the temperature which influences highway conditions (i.e. they will be worse as the temperature decreases);

- the temperature which also influences the number of journeys made, or number of miles travelled (i.e. the miles travelled during Winter, when the weather conditions are bad, will be less because people generally make journeys in spring and summer);

- the bad highway conditions influence driving speed (i.e. people tend to reduce their speed and drive more slowly);

- the driving speed (i.e. in Winter we would not only expect relatively fewer people driving, but also taking more care; implying that we might expect fewer fatal crashes than we would otherwise experience.

The influence of the above factors is presented by the causal BN model in Figure 2.10. In particular, using our understanding of the above factors we can formulate BNs similar to that of Figure 2.10 to combine the statistical information available in a database with other causal subjective factors derived from careful reflection. The objective factors and their relationships are shown with solid lines and arrows in the model example, whereas the subjective factors are shown with dotted lines. Furthermore, the factors introduced interact in a non-liner way, and this helps us to arrive at an explanation for the observed results (i.e. natural causation to drive slower when faced with poor road conditions irrespective of temperature) so the model is able to capture our intuitive beliefs that were contradicted by the counterintuitive results from the simple regression model. Consequently, we might be able to increase the predictive precision of a model, and in return risk assessment and management, by considering subjective information that is important for prediction but which historical data fails to capture.

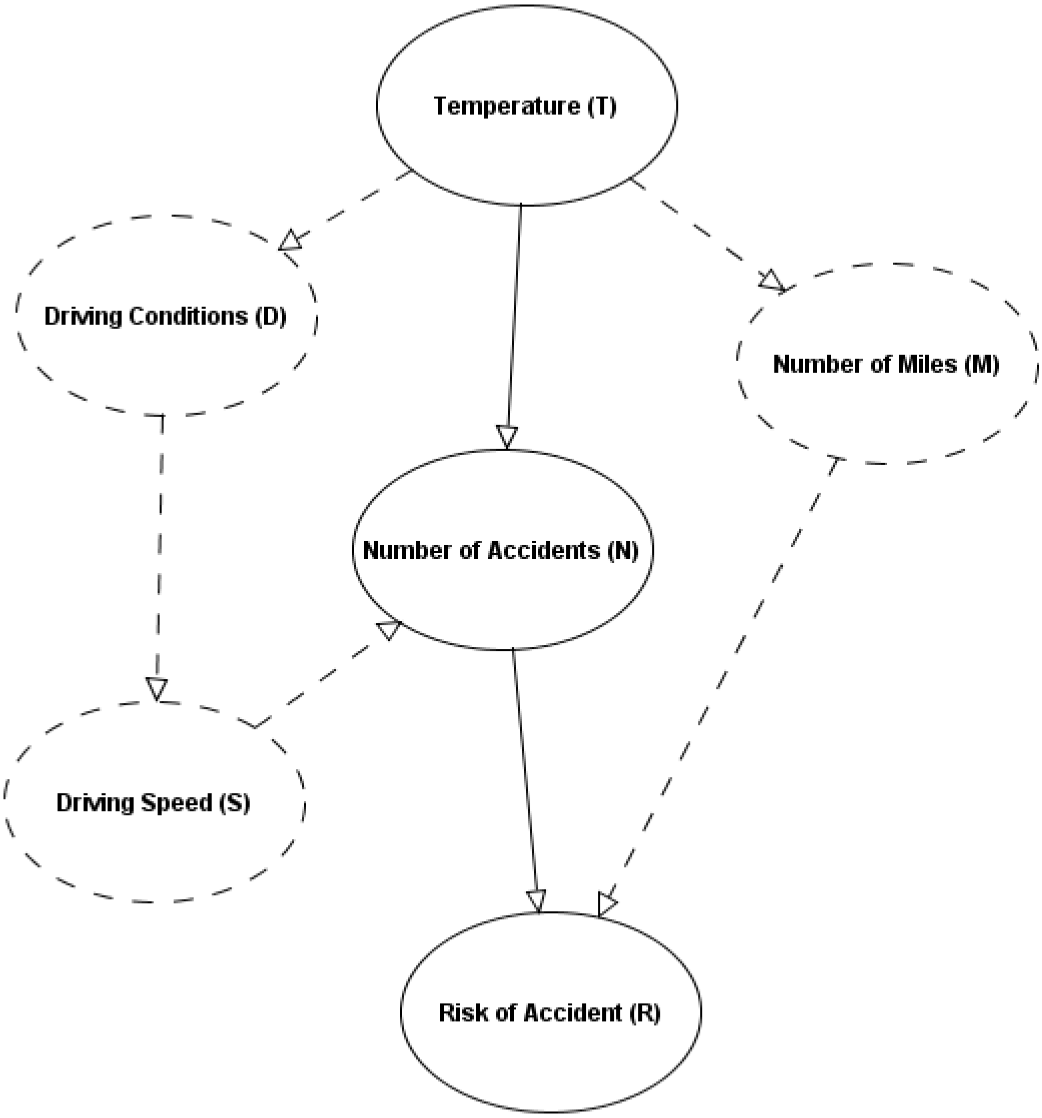


Figure 2.10. Causal BN model for fatal crashes.

2.2.4. Object oriented Bayesian network development

Even though the success of BNs in establishing themselves as an effective and principled framework for knowledge representations and reasoning under uncertainty was generally acknowledged (Pearl, 1988), they were described as being inadequate as a general knowledge representation language for large scale and complex problems (Mahoney & Laskey, 1996). In particular, there are two types of situations in which it becomes inefficient or impractical to model a problem using a single BN (Fenton & Neil, 2012):

a) *When the model contains so many nodes that it becomes conceptually too difficult to understand*; even though one of the key benefits of BNs is their powerful visual aid, it becomes hard to follow when the structure is not simple and contains more than, say, 30 nodes;

b) *When the model contains many similar repeated fragments*; such as the case whereby repeated instances of variable names differ only because of their representation of different point in time.

In an attempt to minimise complexity when constructing large scale models so that we keep them robust, flexible and efficient, (Koller & Pfeffer, 1997) suggested the object oriented Bayesian network (OOBN) approach. Object oriented (OO) design comes from object oriented programming (OOP) and provides a framework for organising abstract data types, whereby *objects* are introduced that consist of data fields and methods together with their interactions, and these objects can be used multiple times (Goldberg & Robson, 1983).

Even though the concept of OOBN is now widely used, none of the BN software tools that are currently available have their inference algorithm implemented in a genuinely OO manner (even though there has been some attempts to formulate this kind of approach (Langseth & Bangso, 2001; Bangso et al., 2003)). In AgenaRisk, an OOBN is simply a BN that is reusable as part of a larger BN. This is achieved with features that allow each BN to have *input* and *output* nodes, and these type of nodes can be managed by an external interface of the BN in order to enable us to link OOBNs together in a well-defined way. For example (Fenton & Neil, 2012), Figure 2.11 presents a rather large model with the possible decompositions indicated, and Figure 2.12 presents the three decomposed OOBNs along with input and output nodes.

Figure 2.11. A rather large model with possible decomposition indicated; based on (Fenton & Neil, 2012).

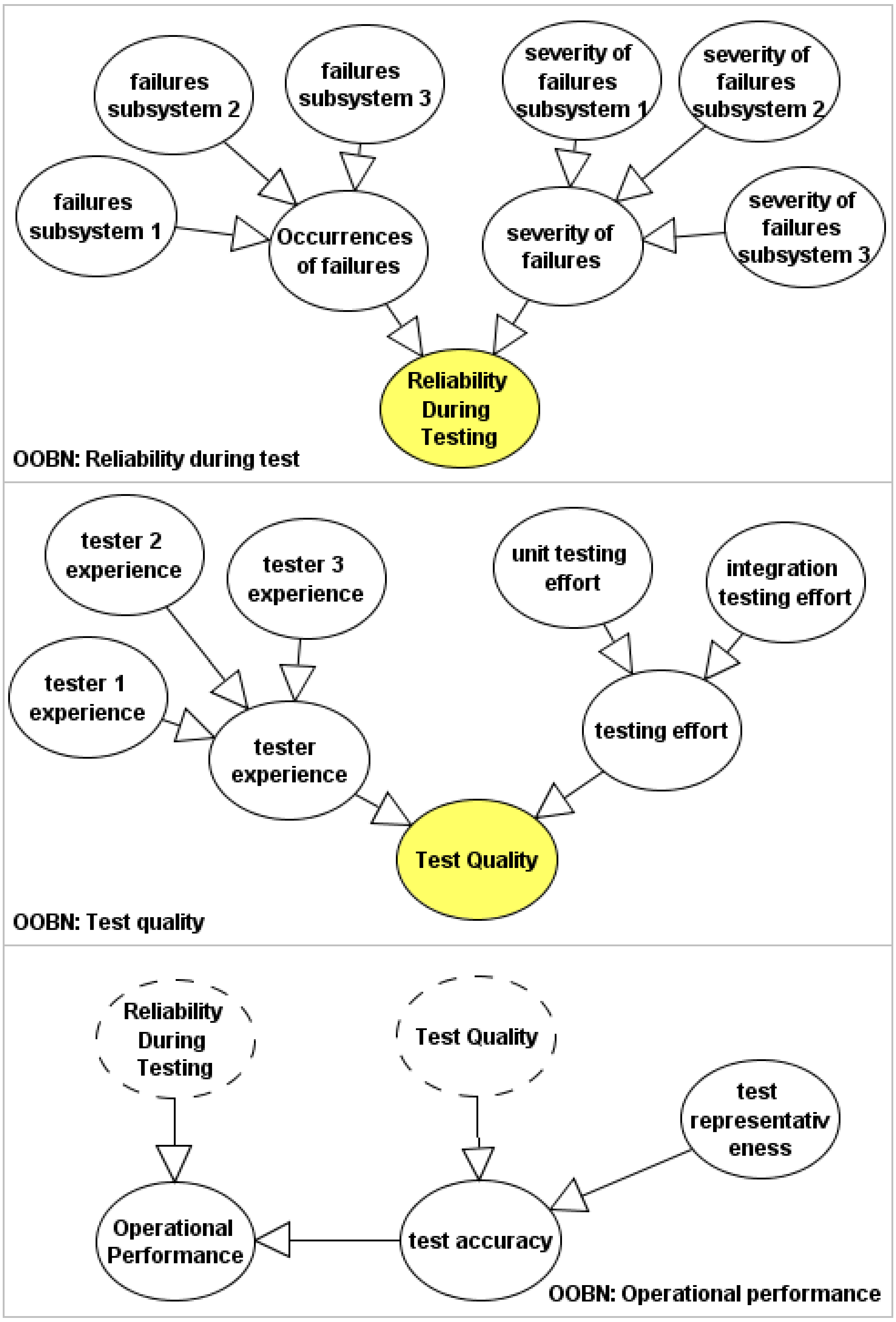


Figure 2.12. The three OOBNs of the larger model presented in Figure 2.11, with input and output nodes[5] based on a three-class decomposition; based on (Fenton & Neil, 2012).

[5] Input nodes are represented by dashed ellipses and have no parent in the class (they correspond to the parameter passed from the associated object). Output nodes are represented by yellow-shaded ellipses and can be parents of nodes outside instances of the class.

2.3 Bayesian vs. frequentist

Comparisons between Bayesian and frequentist approaches for measuring uncertainty have led to an endless debates (Efron, 2005; Vallverdu, 2008), that is beyond the scope of this thesis. In this section we briefly discuss the two approaches and underlining the primary differences between the two. For further reading on measuring uncertainty see (Gigerenzer, 2002; Haigh, 2003).

How Bayesians reason about probability:

- Probability is a measure of a person's degree of belief for an event and thus, subjective information can be incorporated into calculations;
- Subjective beliefs can be assigned to unique events;
- Variables are considered as being uncertain;
- Inference is based on the Bayes' theorem.

How Frequentists reason about probability:

- Probability is a measurable frequency of events and it is determined from repeated experiments;
- Variables are considered as being random;
- Inference is based on the notion of confidence intervals.

2.3.1. The inevitability of subjectivity

Let us assume the following statement (Fenton & Neil, 2012):

> *"There is a 1 in 10 million (or equivalent* 0.0001%*) chance that a meteor will destroy the White House within the next 5 years."*

There is no reasonable frequentist interpretation of the above statement; we cannot run repeated experiments in order to measure the probability of such an event based on the number of times in which the White House is destroyed. However, we can provide a subjective measure of uncertainty based on our current state of knowledge (Fenton & Neil, 2012).

There is no reason why frequentist and subjective approaches cannot work together. Let us assume the following statement (Fenton & Neil, 2012):

> *"There is a* 5% *chance of Spurs winning the FA Cup next year."*

Again to say something about the probability of the above event one has to consider a degree of subjectivity since we are referring to a future event, the FA cup, and there is only one FA Cup next year. We cannot play the identical tournament many times in the same year with the same teams in order to record the frequency of Tottenham winning the tournament. Nevertheless, we can consider the number of times Spurs won the FA Cup in the last few years in order to formulate a prior knowledge of Spur's strength as a team relative to competing adversaries. Of course, in this case past performances are not strong indicators of current performance, but still the frequency of historical FA Cup wins can serve as one of the many factors in predicting this kind of future event.

2.3.2. p-values and significance tests

One of the most common debates involves the null-hypothesis significance testing and p-values, which continue to receive criticism on a consistent basis because they are prone to misinterpretation, they can provide highly misleading evidence against the null hypothesis, they can lead one to reject the null hypothesis when there is really not enough evidence to do so, and many criticise the significant tests for failing to identify genuine differences (Edwards et al., 1963; Berger & Sellke, 1987; Cohen, 1990; Loftus, 1991; Schervish, 1996; Nickerson, 2000; Sterne, 2001; Ziliak & McCloskey, 2004; Cumming, 2005; Doros & Geier, 2005; Ioannidis, 2005a; Ioannidis, 2005b; Killeen, 2005a; Killeen, 2005b; Macdonald, 2005; Wagenmakers & Grunwald, 2005; Armstrong, 2007a; Armstrong, 2007b; Ziliak & McCloskey, 2008). The principle of failing to find evidence that there is a difference does not constitute evidence that there is no difference is described by the statement "Absence of evidence is not evidence of absence" (Douglas & Bland, 1995). After all, Fisher proposed the p-value as an informal measure of evidence against the null hypothesis calling for a combination with other types of evidence for and against that hypothesis (Fisher, 1922; Fisher, 1954).

2.4 Summary

This Chapter introduced the Bayes theorem and demonstrated how it can be used for inference when developing BNs. We briefly discussed the most important limitations we face when we only consider the standard statistical approaches to prediction and risk management, and showed examples of how some of those limitations can be overcome by using BNs and incorporating subjective information. The next Chapter provides a review of the current *state-of-the-art*, within the published academic literature, in football gambling markets and relevant predictive models.

CHAPTER 3

The current *state-of-the-art*

This chapter presents the most comprehensive and up-to-date state-of-the-art on sports prediction models and gambling markets. The different parts of the Chapter correspond to components of the state-of-the-art reviews that appear in the articles listed in Section 1.4.

3.1 Gambling market efficiency

Dixon and Coles (1997) concluded that the UK football betting market is inefficient after a rather simple bivariate Poisson distribution model was able to earn positive returns under specific high-discrepancy trading rules during the English Premier League (EPL) season 1995/96. Similar conclusions have been reported in (Rue & Salvesen, 2000; Kuypers, 2000; Dixon & Pope, 2004). Further, in 2004 (Goddard & Asimakopoulos, 2004) found that the betting market is inefficient at the start[1] and, most notably, at the end of a football season, whereas (Forrest & Simmons, 2008) concluded that published odds appear to be influenced by the number of fans of each club in a match after observing that popular teams are offered more favourable terms on their wagers. Yet, the primary reason why the football betting market is considered by many to be inefficient is perhaps the strong evidence of a favourite-longshot bias (see Chapter 5.2.4) as reported in (Cain et al., 2000, Forrest & Simmons, 2001; 2002).

In contrast to the studies above, other researchers have concluded that the market is efficient. In 1989 (Pope & Peel, 1989) investigated the *ex post* inefficiency of the fixed odds provided between bookmaking firms and concluded that no profitable betting strategies could have been implemented *ex ante* at that time. (Forrest et al., 2005) demonstrated how the efficiency of the market has increased over a five year period with the help of an ordered profit model and showed that their model was unable to make profitable returns against the bookmakers. More recently, Graham and Stott (2008) introduced two forecast models; one based on football results, which is similar to that of (Forrest et al., 2005), and another based on past bookmaking odds in an attempt to compare the bookmaking opinion of various UK teams with the ratings generated by the football results based model. They showed that bookmaking prices were rational and not significantly different than those generated by the model, even though in some cases systematic bookmaking odds biases were observed which could not have been explained. Possibly

[1] This agrees with (Forrest et al., 2005), in which authors showed that over a five-year period, their benchmark statistical model was outperforming bookmaking odds at the very start of the season. However, in all cases the model eventually failed to outperform bookmaker's odds. No claims were made of an inefficient market.

strongest evidence of efficiency are reported in studies in which researchers have attempted to outperform bookmakers' odds by introducing their own forecast models (ranging from very simple to rather sophisticated models), but failed to do so. As a result, other relevant studies have concluded and/or assumed that the betting market is efficient (Peel & Thomas, 1988; 1992; 1997; Vecer et al., 2009).

While this thesis is focused on fixed-odds football betting markets, it is worth noting that there are various other studies within the academic literature which focus on sport betting markets that encompass significant differences in betting behaviour[2]. Discussions regarding such distinct betting markets can be found in (Vergin & Scriabin, 1978; Hausch et al., 1981; Asch et al., 1984; Zuber et al., 1985; Sauer et al., 1988; Thaler & Ziemba, 1988; Golec & Tamarkin, 1991; Shin H., 1991; Shin R. E., 1992; Shin H., 1993; Woodland & Woodland, 1994; Peel & Thomas, 1997; Vaughn Williams & Paton, 1997; Golec & Tamarkin, 1998; Henery, 1999; Jullien & Salanie, 2000; Woodland & Woodland, 2001; Levitt, 2004; Paton & Vaughan Williams, 2005).

[2] Notably, other markets include *pari-mutuel betting* where published odds are determined solely by the behaviour of the bettors (e.g. horse racing), *spread betting* where the returns are based on the accuracy of the bettor (e.g. NFL); *betting exchange* where one bettor can bet against another bettor (e.g. horse racing; this has also recently emerged in UK football betting (betfair, 2000)).

3.2 Football models: Approaches to prediction

While some studies focus on predicting tournament outcomes (Kuonen, 1996; Buchner, et al., 1997; Koning et al., 2003; Halicioglu, 2005a; Halicioglu, 2005b) or league positions (Koning, 2000), our interest is in predicting outcomes of individual matches.

A common approach is the Poisson distribution goal-based data analysis whereby match results are generated by the attack and defence parameters of the two competing teams (Maher, 1982; Dixon & Coles, 1997, Lee 1997; Karlis & Ntzoufras, 2003). A similar version is also reported in (Dixon & Pope, 2004) where the authors demonstrate profitability against the market only at very high levels of discrepancy, but which relies on small quantities of bets against an unspecified bookmaker. A time-varying Poisson distribution version was proposed by (Rue & Salvesen, 2000) in which the authors demonstrate profitability against Intertops (a bookmaker located in Antigua, West Indies), and refinements of this technique were later proposed in (Crowder et al., 2002) which allow for a computationally less demanding model.

In contrast to the Poisson models that predict the number of goals scored and conceded, all other models restrict their predictions to match result, i.e. win, draw, or lose. Typically these are ordered probit regression models that consist of different explanatory variables. For example, (Kuypers, 2000) considered team performance data as well as published bookmakers' odds, whereas (Goddard & Asimakopoulos, 2004; Forrest et al, 2005) considered team quality, recent performance, match significance and geographical distance. (Goddard, 2005) compared goal-driven models with models that only consider match results and concluded that both versions generate similar predictions.

Techniques from the field of machine learning have also been proposed for prediction. In (Tsakonas et. al., 2002) the authors claimed that a genetic programming based technique was superior in predicting football outcomes to other two methods based on fuzzy models and neural networks. More recently, (Rotshtein et al., 2005) claimed that acceptable match simulation results can be obtained by tuning fuzzy rules using parameters of fuzzy-term membership functions and rule weights by a combination of genetic and neural optimisation techniques.

Other studies have considered the impact of specific factors on match outcome. These factors include: home advantage (Clarke & Norman, 1995; Hirotsu & Wright, 2003; Poulter, 2009), ball possession (Hirotsu & Wright, 2003), red cards (Ridder et al., 1994; Vecer et al., 2009)[3], and team form (Knorr-Held, 2000; Hvattum & Arntzen 2010; Leitner et al., 2010).

Recently researchers have considered Bayesian networks and subjective information for football match predictions. In particular, (Joseph et. al., 2006) demonstrated the importance of supplementing data with expert judgement by showing that an expert constructed Bayesian network model was more accurate in generating football match forecasts for matches involving

[3] While this work falls within the scope of our interest, other empirical forecasting studies such as attendance demand (Peel & Thomas, 1989; Peel & Thomas, 1992; Peel & Thomas, 1997; Falter & Perignnon, 2000; Forrest & Simmons, 2002), and the effectiveness of football tipsters (Forrest & Simmons, 2000) do not.

Tottenham Hotspurt than machine learners of MC4, naive Bayes, Bayesian learning and K-nearest neighbour. A model that combined a Bayesian network along with a rule-based reasoner appeared to provide reasonable World Cup forecasts in (Min et al., 2008) through simulating various predifined strategies along with subjective information, whereas in (Baio & Blangiardo, 2010) a hierarchical Bayesian network model that did not incorporate subjective judgments appeared to be inferior in predicting football results when compared to standard Poisson distribution models.

3.3 Using football models to beat the market

While numerous academic papers exist which focus on football match forecasts, only a few of them consider profitability as an assessment tool for determining a model's forecasting capability.

Pope and Peel (1989) 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 (1997) 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 that was based on Maher’s (1982) model. They formulated a simple betting strategy for which the discrepancy of model to bookmakers’ probabilities exceeds a specified level, and showed that the model was only profitable at sufficiently high discrepancy levels. However, at very high discrepancy levels returns were based on as low as 10 sample values; implying that their claims yield high uncertainty since 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"* (Dixon & Coles, 1997). A similar paper by Dixon and Pope (2004) was also published on the basis of 1993-96 data and reported similar results. Rue and Salvesen (2000) 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 (Cain et al., 2000) the authors considered Poisson and Negative Binomial regression models to estimate the number of goals scored by a team in an attempt to exploit the favourite-longshot bias for profitable opportunities, and concluded 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 (2004) proposed an ordered probit regression model to forecast EPL match results in an attempt to test the weak-form efficiency of prices in the fixed-odds betting market. To evaluate the model they considered seasons 1999 and 2000 and even though a loss of −10.5% was reported 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[4]. Forrest et al.

[4] Gross pre-taxed returns of +3.1% and +1.5% for respective seasons beginning 1999 and

(2005) examined the effectiveness of forecasts based on published odds and forecasts generated using a benchmark statistical model with a large number of quantifiable variables relevant to match outcomes. They considered five different bookmaking firms for five consecutive football seasons (1998 to 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. In (Graham & Stott, 2008), the authors 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 (Dixon & Coles, 1997; Dixon & Pope, 2004) and reported negative returns ranging from -2.5% to -15% for all discrepancy levels during seasons 2004 to 2006. In the absence of a particularly successful football model up to that date against market odds, the authors claimed that "*if it was successful, it would not have been published*" (Graham & Stott, 2008). Hvattum and Arntzen (2010) considered the ELO rating system, which was initially developed by (ELO, 1978) for assessing the strength of international chess players, for football match prediction and 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.

2000, and gross returns of $+8\%$ for respective seasons ending 1999 and 2000.

3.4 The value of ratings in terms of forecasting

A rating system provides relative measures of superiority between adversaries. Throughout the football forecasting academic literature, the ability of a football team is most typically dependent on the relevant probabilistic rates of historical match outcomes. Even though there have been numerous attempts in formulating more accurate football forecasting models, the use of pure rating systems has not been extensively evaluated. In fact, only three academic papers appear to have assessed the aid of such systems in football.

In particular, Knorr-Held (2000) appears to be the first to propose a rating system that is primarily intended for rating football teams, even though it is also applicable to other sports. This proposed system was an extended version of the cumulative link model for ordered responses where latent parameters represent the strength of each team. The system was tested according to four different measures and two of them disappointed in performance, whereas an assignment of a team-specific smoothing parameter turned out to be difficult for estimation. In (Hvattum & Arntzen, 2010) the authors suggested the use of the ELO rating for football match predictions (this was initially developed for assessing the strength of chess players (Elo, 1978) and has been adopted to football (Buchdahl, 2003)). They concluded that even though the ratings appeared to be useful in encoding the information of past results for measuring the strength of a team, when used in terms of forecasts it appeared to be considerably less accurate compared to market odds. The ELO rating has also been assessed by (Leitner et al., 2010) along with the FIFA/Coca Cola World ratings (FIFA, 2012) for predicting tournament winners. However, both of these rating systems were said to be clearly inferior to bookmakers' odds, on the basis of EURO 2008 football data, which makes the study consistent with the former.

Harville (1977) stated that a team in American Football should be rewarded for winning *per* se and not for running up the score. Knorr-Held (2000) erroneously assumed that the same logic is applicable to association football on the basis of (Harville, 1977) when formulating performance ratings. In fact, Goddard (2005) demonstrated that no significant difference in forecasting capability is observed between goal-based and result-based regression models for match outcomes in football, and that some advantage is gained by using goal-based (rather than results-based) lagged performance covariates.

3.5 Assessing forecast accuracy

Despite the massive popularity of probabilistic (association) football forecasting models, and the relative simplicity of the outcome of such forecasts (they require only three probability values corresponding to home win, draw, and away win) there is no agreed scoring rule to determine their forecast accuracy. Moreover, the various scoring rules used for validation in previous studies are inadequate since they fail to recognise that football outcomes represent a ranked (ordinal) scale. This raises severe concerns about the validity of conclusions from previous studies. There is a well-established generic scoring rule, the Rank Probability Score (RPS), which has been missed by previous researchers, but which properly assesses football forecasting models.

Defining suitable scoring rules has proven to be extremely difficult, (Murphy & Winkler, 1987; Garthwaite, Kadane, & O'Hagan, 2005), even for apparently ‘simple’ binary forecasts (Jolliffe & Stephenson, 2003), or even when restricted to a specific application domain. For example, in macroeconomics serious concerns over the various scoring rules have been exposed, (Leitch & Tanner, 1991; Armstrong & Collopy, 1992; Hendry, 1997; Fildes & Stekler, 2002). As a result, many have suggested the use of more than one rule in an attempt to obtain an informed picture of the relative merits of the forecasts (Jolliffe & Stephenson, 2003).

We have reviewed all of the previously published studies in which one or more explicit scoring rule was used to evaluate the accuracy of one or more probabilistic football forecasting model. There were nine such studies. The various scoring rules are defined in the Appendix. They fall into two categories:

1. Those which consider only the prediction of the observed outcome (also known as local scoring rules). They are: Geometric Mean, Information Loss, and Maximum Log-Likelihood;

2. Those which consider the prediction of the observed as well as the unobserved outcomes. They are: Brier Score, Quadratic Loss function, and Binary decision.

At least one of the category (a) scoring rules was used in (Dixon & Coles, 1997; Rue & Salvesen, 2000; Hirotsu & Wright, 2003; Goddard, 2005; Karlis & Ntzoufras, 2003; Goddard, 2005; Forrest, Goddard, & Simmons, 2005; Joseph, Fenton, & Neil, 2006; Graham & Stott, 2008; Hvattum & Arntzen, 2010). At least one of the category (b) scoring rules was in (Forrest et al., 2005; Joseph et al., 2006; Hvattum & Arntzen, 2010).

In addition to the above scoring rules, some researchers have proposed and applied different 'ranking' methods of validation, such as the error in cumulative points expected for a team after a number of matches, the RMS and Relative Rank Error of the final football league tables, and pair-wise comparisons between probabilities. At least one of these types of methods has been found in (Pope & Peel, 1988; Dixon & Pope, 2004; Min et al., 2008; Baio & Blangiardo, 2010). However, these methods are beyond the scope of this research study; they do not represent an actual scoring rule, since they cannot provide a measure of accuracy for the prediction of a particular game.

CHAPTER 4

Solving the problem of inadequate scoring rules for assessing probabilistic football forecast models

The novel material introduced in this chapter comes from our publication (Constantinou & Fenton, 2012a), and deals with propriety in the assessment of accuracy of probabilistic football match forecasts.

4.1 Introduction

If a problem has a fixed set of possible outcomes (such as a football match where the outcomes are H, D, A corresponding to Home win, Draw, Away win), a probabilistic forecast model is one that provides predicted probabilities (such as $\{p_H, p_D, p_A\}$ in the case of football) corresponding to the outcomes.

Probabilistic forecasting has become routine in domains as diverse as finance, macroeconomics, sports, medical diagnosis, climate and weather. Some forecasts are conceptually simple (involving a single binary outcome variable) while others are complex (involving multiple possibly related numeric variables). To determine the accuracy of forecast models we use so-called *scoring rules,* which assign a numerical score to each prediction based on how 'close' the probabilities are to the (actual) observed outcome. For a detailed review on the theory of scoring rules and probability assessment in general see (Jolliffe & Stephenson, 2003; Gneiting & Raftery, 2007).

The work in (Gneiting & Raftery, 2007) addresses many of the above problems, by recognising that the underlying measurement scale type of the outcomes for a specific problem should drive the type of scoring rule used. For example, if the problem is to predict a winning lottery number then although the possible outcomes appear to be an ordered set $\{1,2,\ldots,49\}$ the relevant scale type is only *nominal*; if the winning number is 10 then a prediction of 9 is no 'closer' than a prediction of 49 – they are both equally wrong and any scoring rule should capture this. On the other hand, if the problem is to predict tomorrow's temperature in degrees centigrade the relevant scale type is (at least) ordinal (ranked) since if the actual temperature is 10 then a prediction of 9 must be considered closer than a prediction of 49, and any scoring rule should capture this.

Recently, in an attempt to provide a convenient way of constructing scoring rules similar to the Rank Probability Score (RPS), Jose et al. (2009) have commented on football match forecasts by indicating that the outcomes are ordered and that the RPS is suitable for assessing the respective forecasts. The crucial observation we make about football forecasting is that the set of outcomes $\{H, D, A\}$ must be considered as an ordinal scale and not a nominal scale. The outcome D is closer to H than A is to H; if the home team is leading by a single goal then it requires only one goal by the away team to move from H to D. A second goal by the away team is required to move the result on to an A. It follows that if the result is H any scoring rule should penalise the probability assigned to A more heavily than that assigned to D. It turns out that, as obvious as this observation appears, we will show in Section 4.2 that it has been missed in every previous study of football forecasting systems. To demonstrate this we introduce some simple benchmark scenarios along with the result required of any valid scoring rule and show that none of the previously used scoring rules satisfies all of the benchmarks. It follows that all of the previous studies on football forecast models have used inadequate scoring rules. In Section 4.3 we show that the RPS, which is well established standard scoring rule for ordinal scale outcomes, satisfies all the benchmark examples for football forecast models. The implications of this are discussed in Section 4.4.

4.2 Description of previously proposed scoring rules

Below we describe the scoring rules used in the previous studies to assess football match forecasts. In what follows we assume that:

1. In each instance (such as a single football match) there are r possible outcomes (so $r = 3$ for football matches);

2. The model predictions for the outcomes are respective probabilities $\{p_1, p_2, \dots, p_r\}$;

3. The respective actual observed outcomes are $\{e_1, e_2, \dots, e_r\}$. So for football matches the e_is are either 1 or 0 and in such cases the index of the actual observed outcome will be denoted w, so $e_i = 1$ if $i = w$ and 0 if $i \neq w$.

A scoring rule is actually defined in terms of two components:

1. A score for an individual instance given a forecast for that instance (e.g. a single football match);

2. A method for defining the cumulative scores over a set of instances. With the exception of the geometric mean, the method used is either the arithmetic mean of the individual scores or the total of the individual score over those instances. The geometric mean, in contrast uses a multiplicative function for the cumulative score; meaning that it punishes individual bad predictions heavier.

For the purposes of this research study it is sufficient to consider only how the scoring rule is defined for individual instances. These definitions are:

1. **Binary Decision**: The score is 1 if $p_w > p_i$ for each $i \neq w$ and 0 otherwise. The Binary Decision rule takes into consideration all of the probabilistic values since it seeks the highest one for assessment. However, the rule does not generate a score based on the values;

2. **Brier Score**: also known as Quadratic Loss, the Brier Score (Brier, 1950) is

$$\sum_{i=1}^{r}(p_i - e_i)^2$$

3. **Geometric Mean:** For an individual instance the score is simply p_w;

4. **Information Loss**: For an individual instance this is defined as $-log_2 p_w$. In order to avoid the *zero-frequency* problem (whereby the informational loss is minus infinity when p_w is zero) it is expected that non-zero probabilities are assigned to every outcome.;

5. **Maximum Log-Likelihood Estimation (MLLE)**: Maximum likelihood estimation is known as an approach to parameter estimation and inference in statistics, which states that the desired probability distribution is the one that makes the observed data 'most likely'. Informational Loss and Maximum Log-Likelihood estimation differ in equation but generate identical forecast assessment. As in most cases, we present the MLLE over the MLE since it generates identical assessment while reducing the computational cost significantly (Myung, 2003). For a likelihood test, the Binomial distribution with parameters

n (trials), and s (successes) could be used. However, since we only have one observation for every match prediction (n and s are equal to 1), the Log-likelihood is $ln(p_w)$. Therefore, in order to avoid unnecessary calculations, we simply define the MLLE to be $ln(p_w)$ for an individual match.

The work in (Gneiting & Raftery, 2007) demonstrates that none of the above scoring rules are suitable for a problem whose outcomes are on a ranked scale, and hence they are unsuitable for assessing football forecast models. We demonstrate this informally by introducing five 'benchmark' scenarios (Table 4.1) in each of which we have a match that has two 'competing' prediction models α and β together with an actual outcome. The proposed scenarios demonstrate the limitations introduced by scoring rules that do not consider the distribution of probabilities as an ordinal scale. Many additional different scenarios could have been proposed, but these ones are representative of actual football match forecasts and also cover a wide range from certainty that one team will win through to a case where equal probabilities are assigned to both teams[1]. In each case it is clear intuitively which of the predictions should be scored higher.

[1] We also exploit the symmetry of the problem by only considering scenarios in which the same team (Home as opposed to Away) is favoured. For example, the problem associated with comparing the two predictions in Match 1 is the same as the problem of comparing the predictions $\{0, 0, 1\}$ and $\{0, 0.1, 0.9\}$.

Table 4.1. Hypothetical forecasts by models α and β, and results for Matches 1 to 5.

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		

This is because:

a) Match 1: (Taking account of perfect accuracy) Model α predicts the actual outcome with total certainty and hence must score better than any other, less perfect, predicted outcome;

b) Match 2: (Taking account of predicted value of the observed outcome) Both models α and β assign the highest probability to the winning outcome H, with the remaining two outcomes evenly distributed. Since the observed value of α is higher than that of β, it must score higher;

c) Match 3: (Taking account of distribution of the unobserved outcomes) Given that the observed outcome here is D, both of the unobserved outcomes are equally distanced from the observed one. Hence, the ordering concern here is eliminated. Still, a scoring rule must identify that model α is more accurate since its overall distribution of probabilities is more indicative of a draw than that of β (which strongly predicts a home win);

d) Match 4: (Taking account of ordering when the set of unobserved outcomes are equal) Both models α and β assign the same probability to the winning outcome H. This time, however, they also assign the same probability values (but in a different order) to the unobserved outcomes (0.25 and 0.15). But, a scoring rule must identify that model α is more accurate since its overall distribution of probabilities is more indicative of a home win;

e) Match 5: (Taking account of overall distribution) Although α predicts the actual outcome H with a lower probability than β the distribution of α is more indicative of a home win than β. This match is the most controversial, but it is easily explained by considering a gambler who is confident that the home team will not lose, and so seeks a *lay* bet (meaning explicitly that the bet wins if the outcome is H or D). Assuming that α and β are forecasts presented by two different bookmakers, bookmaker α will pay less for the winning bet (this bookmaker considers that there is only 10% probability the home team will lose, as opposed to bookmaker β who considers it a 20% probability).

Table 4.2 presents the results of the previously used football scoring rules for the benchmark scenarios and determines the extent to which they satisfy the benchmark for each of those forecasts presented in Table 4.1. A tick means that the scoring rule correctly scores model α higher than β. A double cross means that the scoring rule incorrectly scores model β higher than α. A single cross means that the scoring rule returns the same value for both models. Where necessary, the score is rounded to 4 decimal points. For the rules Binary Decision, Geometric Mean and MLLE, a higher score indicates a better forecast; whereas for the rules Brier Score and the Information Loss a lower score indicates a better forecast.

Table 4.2. Applying the specified scoring rules to each benchmark presented in Table 4.1

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

None of the scoring rules returns the 'correct' outcome for all 5 scenarios. Indeed, all of the scoring rules fail to correctly identify model α as superior for scenarios 4 and 5.

4.3 The Rank Probability Score

The RPS was introduced in 1969 (Epstein, 1969). It is both *strictly proper*[2] (Murphy, 1969) and *sensitive to distance*[3] (Murphy, 1970). The RPS has been described as a particularly appropriate scoring rule for evaluating probability forecasts of ordered variables (Murphy, 1970). In general the RPS for a single problem instance is defined as:

$$RPS = \frac{1}{r-1}\sum_{i=1}^{r-1}\left(\sum_{j=1}^{i}(p_j - e_j)\right)^2$$

where r is the number of potential outcomes, and p_J and e_J are the forecasts and observed outcomes at position j. The RPS represents the difference between the cumulative distributions of forecasts and observations, and the score is subject to a negative bias that is strongest for small ensemble size (Jolliffe & Stephenson, 2003). Since the scoring rule is sensitive to distance, the score penalty increases the more the cumulative distribution forecasted differs from the actual outcome (Wilks, 1995). For a detailed analysis on the RPS see (Epstein, 1969).

Table 4.3 presents the generated score for each of the scenarios presented in Table 4.1, along with the respective cumulative distributions (forecasted and observed). A lower score (rounded to 4 decimal points) indicates a better forecast. Unlike the previous metrics the RPS correctly scores α as 'best' for all 5 matches.

---

[2] A scoring rule is *strictly proper* if it is uniquely optimised by the true probabilities.

[3] A scoring rule is *sensitive to distance* if it takes into account the ordering of events.

Table 4.3. Score generated by the RPS for each hypothetical forecast presented in Table 4.1, along with the respective cumulative distributions (forecasted and observed).

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

It is important to note that there is a possible debate about the RPS (and also the Brier score) in relation to the Match 3 forecast scenarios. For both the Brier score and RPS, the squared measurement of probabilities results in scores that are higher (worse forecasts) for unobserved outcomes which are unevenly distributed. Hence, the Brier score and the RPS are, respectively, the only rules which determine model α as the best model for the particular forecast. We feel this is a clear strength of the RPS - all of the football experts we informally sampled identified model α as the best for this scenario.

4.4 Implications and conclusions

Measuring the accuracy of any forecasting model is a critical part of its validation. In the absence of an agreed and appropriate type of scoring rule it might be difficult to reach a consensus about: a) whether a particular model is sufficiently accurate; and b) which of two or more competing models is 'best'. In this study, the fundamental concern is the inappropriate assessment of forecast accuracy in association football, which may lead in inconsistencies, whereby one scoring rule might conclude that model α is more accurate than model β, whereas another may conclude the opposite. In such situations the selection of the scoring rule can be as important as the development of the forecasting model itself, since the score generated practically judges the performance of that model. On the one hand, an outstanding model might be erroneously rejected while on the other hand a poor model might be erroneously judged as acceptable.

We have shown that, by failing to recognise that football outcomes are on an ordinal scale, all of the various scoring rules that have previously been used to assess the forecast accuracy of football models are inadequate. They fail to correctly determine the more accurate forecast in circumstances illustrated by the benchmark scenarios of Table 4.1. This failure raises serious concerns about the validity and conclusions from previous studies that have evaluated football forecasting models. What makes the failure of all previous studies to use a valid scoring rule especially surprising is that there was already available (before any of the studies were conducted) a well-established scoring rule, the RPS, that avoids the inconsistencies we have demonstrated.

With the relentless increase in interest in football forecasting it will become more important than ever that effective scoring rules for forecast models are used. Although we are not suggesting that the RPS (or the alternative proposition) is the only valid candidate for such a scoring rule, we have shown that (unlike the previous scoring rules used) it does satisfy the basic benchmark criteria expected.

Given the massive surge in popularity of the sport and its increasing dominance in sport betting internationally, it is important to note that we have only considered the assessment of forecast accuracy and not profitability. We cannot claim that a forecasting model assessed as more accurate than a bookmaker by such a rule will necessarily indicate profitability. After all, profit is not only dependent on the accuracy of a model but also on the specified betting methodology. Other researchers have already concluded that there is a weak relationship between different summary error statistics and profit measures (Leitch & Tanner, 1991), whereas others have concluded that it might be best to combine profitability methodologies with a proper forecast assessment rule (Wing et al., 2007). Yet, it is evident that profitability is dependent on accuracy and not the other way around. Accordingly, higher forecast accuracy indicates a higher prospective profit which denotes the importance of propriety in forecast assessment.

CHAPTER 5

Evidence of an (intended) inefficient Association Football gambling market

The novel material introduced in this chapter comes from our paper submitted for publication (Constantinou & Fenton, 2012b), and evaluates the efficiency of the football gambling market based on their published odds.

5.1 Introduction

A gambling market is usually 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. This chapter evaluates the efficiency of the football betting market by taking into consideration 11 years of relevant information. In contrast to earlier studies, we primarily show that:

1. the accuracy between bookmakers is extremely consistent and bookmaking accuracy has not improved over the last decade;

2. profit margins have been dramatically reduced over the last decade and can be statistically significant between bookmakers; implying that the published odds of one bookmaker cannot be considered as representative of the overall market;

3. profit margins per distinct match can be significant even when considering only one bookmaker and one football division;

4. some arbitrage opportunities are found between the odds offered by a small number of bookmaking firms;

5. both systematic and significant adjustments of published odds occur at least daily. In many cases the changes cannot be explained by rational qualitative factors and hence may be due to betting volumes.

We conclude that the football betting market is deliberately inefficient in an attempt to accomplish commercial objectives but that this inefficiency can only be exploited by a very limited number of bettors.

The chapter is structured as follows: Section 5.2 reports on bookmakers' accuracy of the odds and consistent biases; Section 5.3 reports on bookmakers' introduced profit margins and positive arbitrage opportunities;

Section 5.4 demonstrates how published odds are adjusted over time and attempts to explain the rationality behind this behaviour; Section 5.5 discusses the results along with the implications, and we provide our conclusions along with potential future work in Section 5.6.

5.2 Bookmakers' accuracy

Since the bookmakers increase profitability by encouraging bettors to place as many bets as possible, their profit is not only determined by the introduced profit margin (see Section 5.3), but also by the accuracy of their published odds which should, therefore, represent a good approximation of the 'true' probabilities of any particular match without introducing biases. In this section we examine the degree of variation between bookmakers with regards to the accuracy of the normalised[1] odds and we report on the difference in such odds for different various football leagues. We show that there has been no change in the accuracy of published odds over the last decade, and we illustrate consistent biases which exist in published odds.

5.2.1. Data and methodology

The data used for this study is available at www.football-data.com. For forecast assessment we make use of the Rank Probability Score (RPS). We explained why it was the most rational scoring rule of those that have been proposed and used for football outcomes in Chapter 4.

[1] The odds are normalised such that the profit margin (see Section 5.3) is eliminated and the sum of the probabilities over the possible events is equal to 1.

5.2.2. The accuracy of the odds per football league

We have evaluated the accuracy of bookmakers' odds for different leagues and divisions. So this section can be seen as 'league accuracy' rather than bookmakers' accuracy. For this task, we have selected four top division leagues; those from England and Spain which serve as the two most popular and (currently) strongest top division leagues, and those from Belgium and Greece which serve as two considerably less popular and weaker leagues than the former. We also include another five non-top division leagues from England and Spain. The purpose was to test how the accuracy of the odds may differ for top division leagues from different countries, and for different levels of divisions within the same country.

Table 5.1 presents the accuracy scores of the William Hill odds for seasons 2000 to 2011, for all of the nine football leagues described above. The leagues are separated by season, country and division. For simplicity and an easier interpretation of the divisions for each country, in this section we assume that division 1 is the top division for every country (e.g. instead of referring to the Premier League for England), and each subsequent lower league receives an increment of one.

The results appear to be rather surprising. To begin with, the mean accuracy scores from Table 5.1 shows no indications of an improved forecast performance over a period of 11 years, as many have intuitively assumed or concluded (see Section 5.5). The column ‘Mean’ summarises the mean accuracy per season and demonstrates that the bookmaking performance has been incredibly consistent over the last decade, and this is also true for each of the 9 distinct leagues. This suggests that a) apparently bookmakers have failed to improve their forecast performance and b) a richer historical football database does not necessarily imply higher accuracy.

Table 5.1. RPS assessment of William Hill odds per specified football season, football division and country. Smaller score indicates greater accuracy.

Season	England Div. 1	Spain Div. 1	Belgium Div. 1	Greece Div. 1	England Div. 2	Spain Div. 2	England Div. 3	England Div. 4	England Div. 5	Mean
2000/01	0.2021	0.1982	0.1894	-	0.2128	-	0.2139	0.2035	-	0.2033
2001/02	0.1985	0.2076	0.1971	-	0.2135	0.2225	0.2080	0.2111	-	0.2083
2002/03	0.2032	0.2104	0.2044	0.1837	0.2194	0.2118	0.2145	0.2187	-	0.2083
2003/04	0.2033	0.2169	0.1949	0.1562	0.2156	0.2097	0.2099	0.2150	-	0.2027
2004/05	0.1927	0.1962	0.1939	0.1525	0.2145	0.2218	0.2186	0.2175	-	0.2010
2005/06	0.1952	0.2101	0.2044	0.1671	0.2088	0.2170	0.2215	0.2184	0.2156	0.2065
2006/07	0.1953	0.2103	0.1894	0.1869	0.2223	0.2182	0.2213	0.2190	0.2258	0.2098
2007/08	0.1799	0.2196	0.2060	0.1841	0.2182	0.2050	0.2134	0.2312	0.2134	0.2079
2008/09	0.1914	0.2037	0.2025	0.1785	0.2141	0.2008	0.2198	0.2163	0.2118	0.2043
2009/10	0.1832	0.1817	0.2045	0.1795	0.2081	0.2090	0.2012	0.2232	0.2099	0.2000
2010/11	0.2002	0.1908	0.1919	0.2007	0.2208	0.2103	0.2205	0.2172	0.2061	0.2065
Mean	0.1950	0.2041	0.1980	0.1766	0.2153	0.2126	0.2148	0.2174	0.2138	

Further, one might intuitively expect that bookmakers pay more attention to higher popularity leagues due to the larger number of bets they expect to receive and thus, the generated odds might be more accurate when compared to other less popular leagues. However, results show that this is not exactly the case when it comes to accuracy. Although results suggest that the overall bookmaking accuracy for top division leagues is consistently higher than lower division leagues, the accuracy does not continue to diminish while further moving to weaker divisions. Moreover, the accuracy is mostly dependent on the predictability[2] of the league rather than its popularity since the odds provided to the Greek league (which is less popular than the English

[2] A football league L_1 is more predictable than another league L_2 when matches played within L_1 result into less 'surprises' than those of L_2. As a result, the team-ranking of L_1 is likely to be more consistent after each consecutive season than the team-ranking of L_2.

and Spanish leagues, but more 'predictable') were consistently more accurate than any other league.

5.2.3. The accuracy of the odds per bookmaker

The above results were dependent on the odds provided by a single bookmaker, William Hill. In contrast to (Pope & Peel, 1989), recent studies have concluded and/or assumed that little information is lost by concentrating on just one bookmaker (Forrest & Simmons, 2002; Forrest et al., 2005). We test this notion by comparing the accuracy of the odds provided by seven different bookmakers for the top English division from period 2007 to 2011[3]. Table 5.2 reports on the summary statistics.

Without further tests, it should be obvious by looking at the summary statistics that the accuracy of the normalised odds between bookmakers is extremely consistent, and the difference in the odds per season is much greater than the difference in the odds between bookmakers for the same season.

Table 5.2. Summary statistics of RPS assessment of normalised published odds per specified bookmaker and EPL season.

Bookmaker	RPS Mean for seasons:				RPS Median for seasons				S.D. of RPS for seasons:			
	07/08	08/09	09/10	10/11	07/08	08/09	09/10	10/11	07/08	08/09	09/10	10/11
William Hill	0.1799	0.1914	0.1832	0.2002	0.1508	0.1617	0.1476	0.1662	0.1130	0.1202	0.1331	0.1307
BET365	0.1769	0.1915	0.1824	0.2003	0.1447	0.1609	0.1489	0.1678	0.1168	0.1226	0.1337	0.1348
Bwin	0.1782	0.1921	0.1843	0.2010	0.1467	0.1565	0.1537	0.1659	0.1117	0.1194	0.1324	0.1340
Gamebookers	0.1777	0.1918	0.1833	0.2014	0.1455	0.1585	0.1497	0.1679	0.1118	0.1173	0.1295	0.1310
Interwetten	0.1798	0.1916	0.1832	0.2008	0.1515	0.1606	0.1471	0.1661	0.1045	0.1123	0.1270	0.1248
Ladbrokes	0.1799	0.1927	0.1846	0.2004	0.1515	0.1604	0.1528	0.1681	0.1075	0.1158	0.1323	0.1328
Sportingbet	0.1786	0.1921	0.1836	0.2006	0.1482	0.1620	0.1482	0.1675	0.1099	0.1164	0.1288	0.1304

[3] Relevant information previous to 2007 was not available for all of the 7 bookmakers.

Table 5.2 provides further evidence of the phenomenon of predictability as discussed earlier in Section 5.2.2. In particular, the accuracy results suggest that in certain seasons the teams perform as expected. For example, for a decade up until the EPL season 2007/08 the same four teams (Manchester United, Chelsea, Arsenal and Liverpool) not only consistently dominated the top four positions (which guarantee Champions League spots) but were also some distance ahead of the remaining teams. Thereafter, this dominance was challenged by Tottenham and Man City who respectively came from nowhere to claim a top 4 spot at the expense of Liverpool.

This demonstrates a phenomenon identical to that observed in Table 5.1 between the Greek league and the rest. However, in this case we observe this phenomenon for the same league, but for different seasons, which is consistent with our claims on predictability; but which also suggests the limitation of the data-only approaches to prediction.

5.2.4. The favourite-longshot bias

In gambling markets, the favourite-longshot bias refers to the preference of the bettor in backing risky outcomes, which are also referred to as longshots. For example, consider a game between a top team who plays at their ground against a very weak team. Under such scenario, the odds for a home win are approximately 1.10 and apparently, placing a £100 bet to win only £10 is not in the standard bettor's best interest. It seems that bookmakers take advantage of this behaviour and publish odds which are biased against the bettors. In particular, bookmakers are believed to exploit this behaviour and increase profitability by offering more-than-fair odds for 'safe' outcomes, and less-than-fair odds for 'risky' outcomes. This phenomenon is observed in many different markets (Ali M. , 1977; Quandt, 1986; Thaler & Ziemba, 1988; Shin H., 1991, Shin R. E., 1992; Shin H., 1993; Woodland & Woodland, 1994; Vaughn Williams & Paton, 1997; Golec & Tamarkin, 1998; Jullien & Salanie, 2000), and various theories exist, such as risk-loving behaviour, on why people are willing to bet on such uncertain propositions (Sobel & Raines, 2003; Snowberg, 2010).

In order to investigate the degree of this bias we simulate bets for each potential betting outcome that follows the standard form of a football match $\{p(H), p(D), p(A)\}$ and record the resulting cumulative returns. For the purpose of our analysis $p(A)$ is considered to be the longshot (since away wins are far less frequent than home wins). In particular, Figure 5.1 demonstrates the cumulative returns after simulating one-pound bets on all of the three outcomes, for 380 matches per season, against the prices offered by William Hill. We have used seven years of William Hill odds to examine this phenomenon for the matches in the EPL and for seasons 2004/05 to 2010/11. Each figure represents a distinct season, and the cumulative returns of each outcome are illustrated by the three different lines for each graph. As in previous studies, the results illustrate strong evidence of the favourite-longshot bias. In 6 out of the 7 seasons examined, the odds assigned to away teams appear to generate noticeably lower cumulative returns that those observed by the remaining two outcomes. Indeed, in many cases the cumulative returns result in a loss for the bookmakers for the outcomes $p(H)$ and $p(D)$. Clearly, this phenomenon still exists and it is extremely consistent over a period of seven seasons.

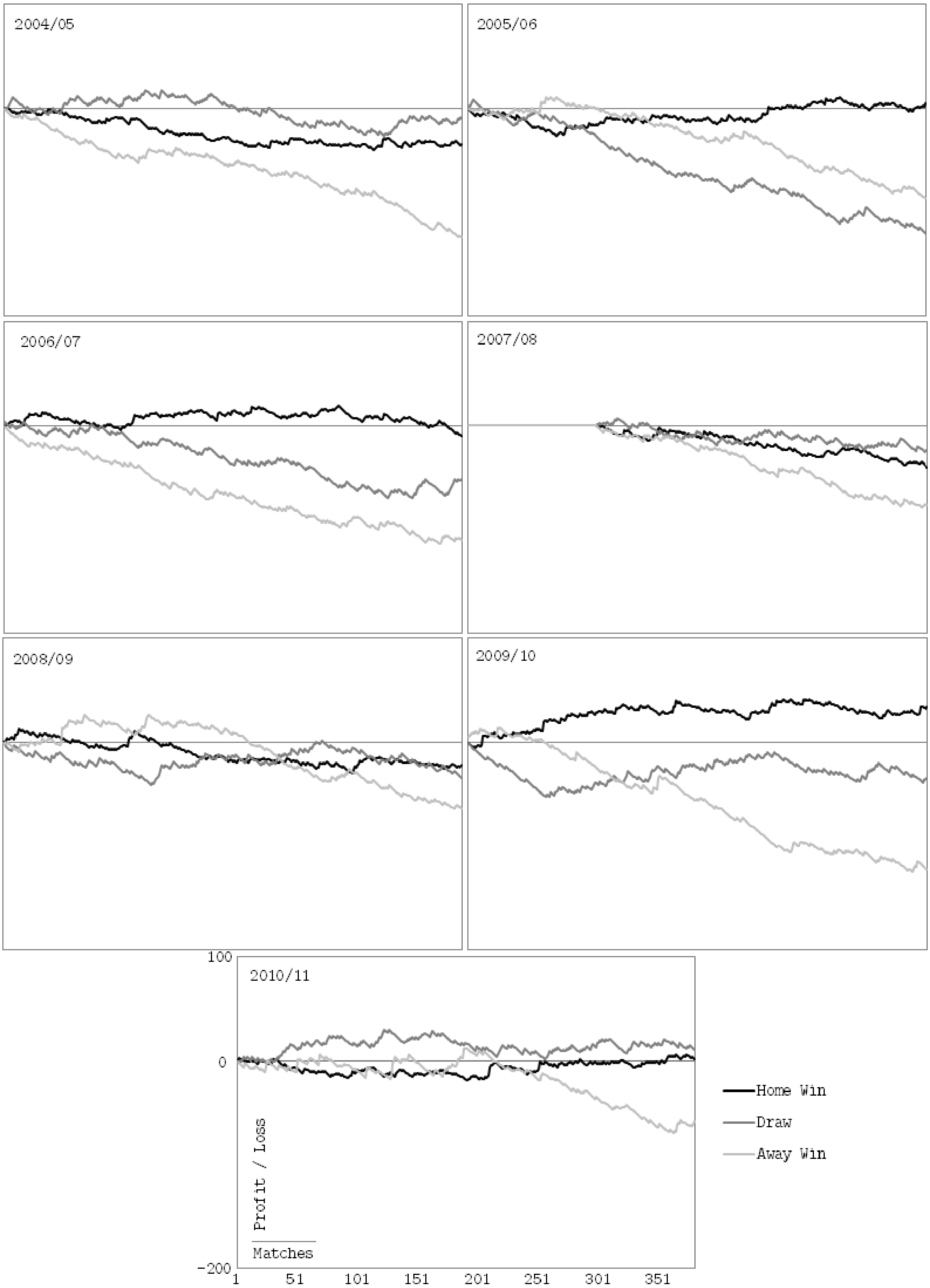


Figure 5.1. Cumulative returns after simulating a £1 bet on William Hill odds over a period of seven years for the EPL matches (season 2007/08 ignores the first few weeks due to the unavailability of the odds).

5.3 Profit margins

The bookmakers' profit margin, also known as "over-round", refers to the margin by which the sum of the probability odds of the total outcomes exceeds 1 (thus, making the odds unfair for bettors). A lower profit margin results in less-unfair published odds. In short, the profit margin indicates the precise profit a bookmaker receives if bets are distributed such that the bookmaker pays the same amount of winnings whatever the outcome of the match. Since it is almost impossible that the bets are distributed as specified above (as discussed, the favourite-longshot bias ensures it is not), the profit margin is just an approximation of the average profit expected.

5.3.1. Profit margins introduced per football league

Similar to the previous section, we follow the same procedure and make use of identical data. Table 5.3 presents the observed profit margin of identical leagues and football season to those reported earlier in Table 5.1. The results reveal a steadily decreasing profit margin. Yet, the observed reduction is only significant over the last 3 or 4 latest seasons. It is a rather interesting fact that the diminished profit margin for English divisions 2 to 4 is lower than top divisions in Belgium and Greece.

Table 5.3. Profit margin introduced in William Hill odds per specified football season, football division and country.

Season	England League 1	Spain League 1	Belgium League 1	Greece League 1	England League 2	Spain League 2	England League 3	England League 4	England League 5	Mean
2000/01	12.53%	12.55%	13.76%	-	12.56%	-	11.65%	11.69%	-	13.32%
2001/02	12.52%	12.51%	13.66%	-	12.53%	13.64%	11.65%	11.68%	-	14.69%
2002/03	12.51%	12.56%	13.44%	13.61%	12.50%	13.68%	12.50%	12.52%	-	15.32%
2003/04	12.49%	12.50%	13.22%	13.58%	12.50%	13.39%	12.49%	12.49%	-	14.82%
2004/05	12.49%	12.35%	12.58%	12.42%	12.48%	12.77%	12.47%	12.43%	-	14.88%
2005/06	12.49%	12.46%	12.46%	12.46%	12.46%	12.53%	12.45%	12.47%	12.57%	14.86%
2006/07	12.49%	12.47%	12.43%	12.48%	12.48%	12.45%	12.44%	12.44%	12.64%	14.93%
2007/08	12.37%	12.37%	12.38%	12.45%	12.33%	12.44%	12.44%	12.40%	12.42%	14.87%
2008/09	7.01%	11.40%	12.34%	11.70%	11.04%	12.45%	11.12%	11.02%	12.23%	13.56%
2009/10	7.35%	10.10%	10.28%	9.39%	11.21%	9.91%	11.48%	11.57%	12.49%	13.72%
2010/11	6.50%	6.68%	9,35%	7.10%	6.75%	9.65%	6.80%	6.76%	10.25%	11.56%
Mean	10.98%	11.63%	12.35%	11.69%	11.71%	12.29%	11.59%	11.59%	12.10%	

5.3.2. Profit margins introduced per bookmaker

In Section 5.2.3 we have showed that the accuracy of the normalised odds of one bookmaker can be representative of any bookmaker due to their extreme consistency. In this section we perform the same test before normalisation and we compare the profit margins introduced by the seven bookmakers.

Table 5.4 presents the mean profit margin introduced in the EPL per specified bookmaker over a period of four years. Unlike normalised accuracy, the results here show that the introduced profit margins can be significantly different per bookmaker; implying that the published odds of one bookmakers cannot be representative of the whole market.

Indeed, Table 5.4 reveals that a) the introduced profit margins can have more that 100% difference per bookmaker for the same league and season, and b) it is most likely that bookmakers will decrease their profit margin after each consecutive year due to competitiveness; yet some bookmakers may still decide to keep their introduced profit margins constant over a number of successive years, whereas others may even introduce an increase.

Table 5.4. Mean profit margins introduced per specified bookmaker and EPL season.

Bookmaker	Season 2007/08	Season 2008/09	Season 2009/10	Season 2010/11
William Hill	12.37%	7.01%	7.35%	6.50%
BET365	5.98%	5.31%	5.43%	5.44%
Bwin	10.06%	10.07%	8.30%	8.01%
Gamebookers	7.45%	7.29%	7.75%	7.68%
Interwetten	11.39%	10.21%	8.36%	10.13%
Ladbrokes	12.19%	9.26%	7.48%	6.49%
Sportingbet	10.13%	10.14%	10.12%	10.12%

In view of the above evidence, we have decided to further investigate how each bookmaker behaves against each of the distinct matches over a whole season. Figures 5.2 and 5.3 illustrate the profit margin introduced per bookmaker for each successive EPL match during season 2010/11. Four out of the seven bookmakers, BET365, Gamebookers, Interwetten and Sportingbet appear to provide a rather consistent profit margin for successive matches throughout the whole season. In contrast, bookmakers William Hill, bwin and Ladbrokes demonstrate significant fluctuations. What is even more interesting is that the observed fluctuations introduced by the three specified bookmakers are dissimilar (as illustrated in Figures 5.2 and 5.3). In particular, William Hill introduced roughly double profit margins during the last gameweek[4] of the season (Figure 5.2a), bwin introduced a significantly diminished profit margin in 11 (out of 380) matches (Figure 5.2c), and Ladbrokes introduced a significantly raised profit margin in 6 matches (Figure 5.3f).

[4] The top English division has 38 gameweeks, and normally 10 matches for each gameweek for a total of 380.

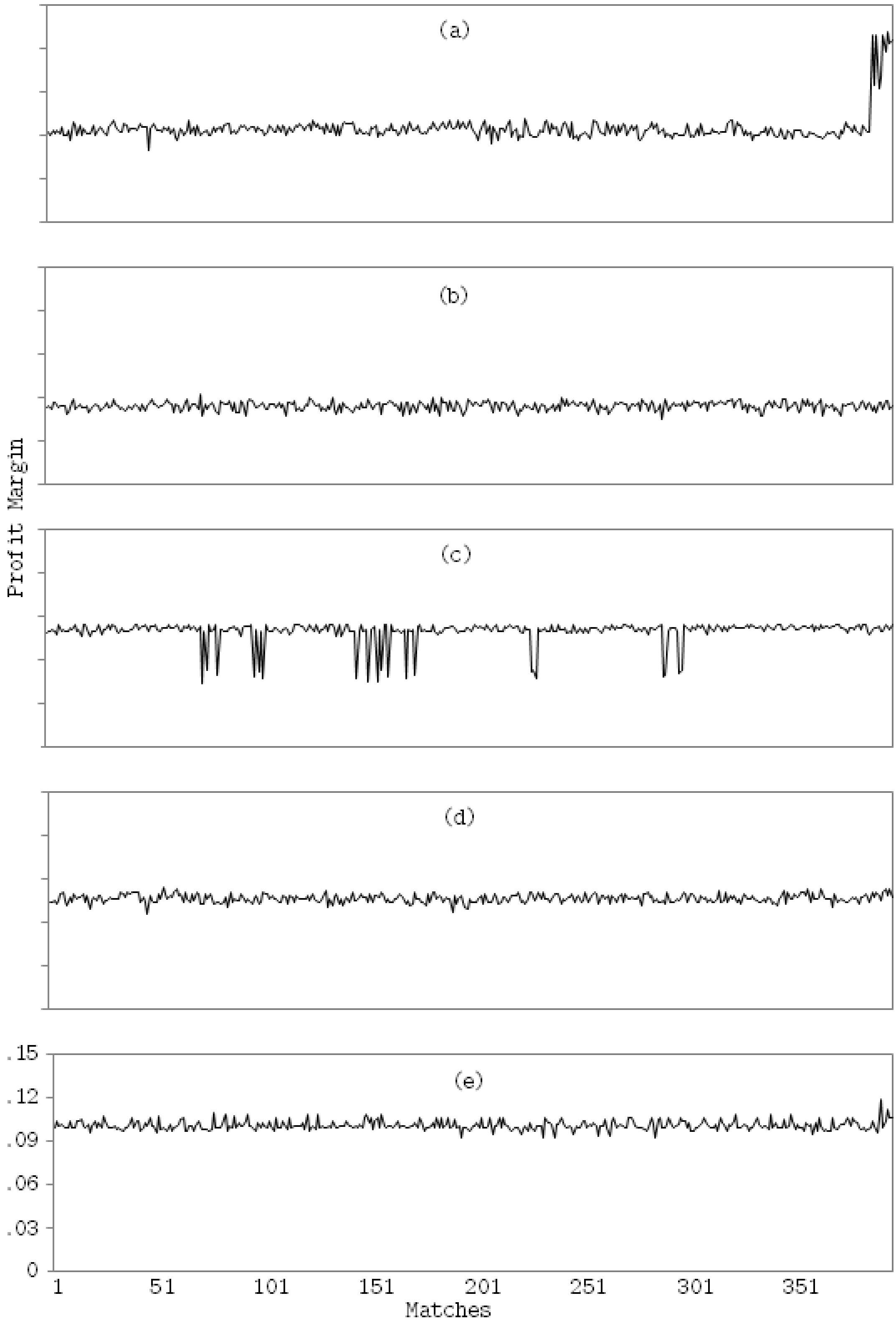


Figure 5.2. Successive profit margins introduced per bookmaker for football matches during the EPL season 2010/11, where (a) is William Hill, (b) is BET365, (c) is bwin, (d) is Gamebrookers and (e) is Interwetten.

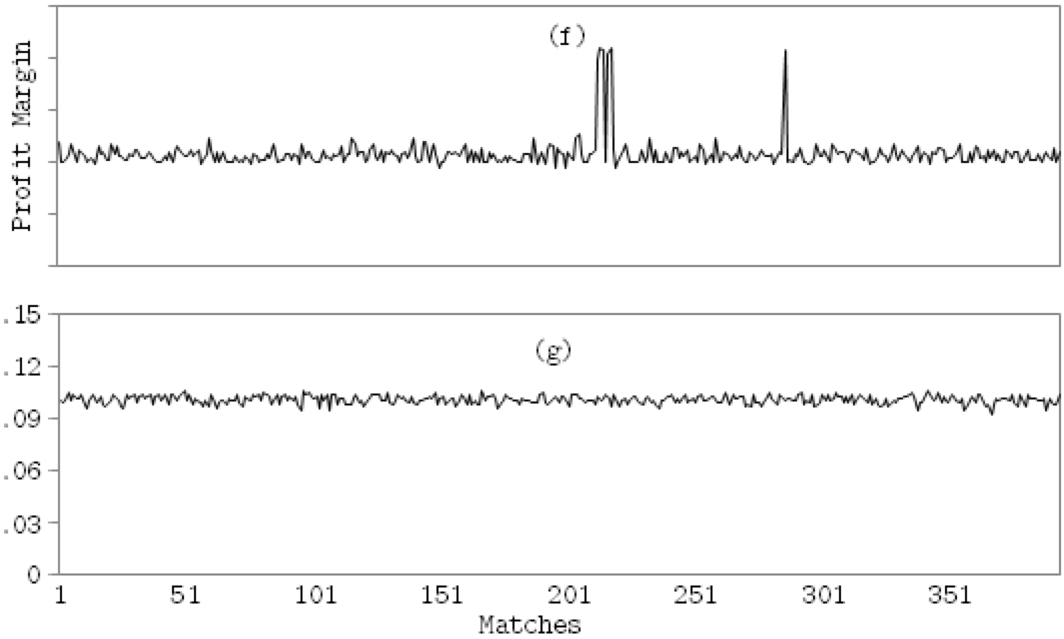


Figure 5.3. Successive profit margins introduced per bookmaker for football matches during the EPL season 2010/11, where (f) is Ladbrokes and (g) is Sportingbet.

If we assume that during that particular season the teams Man United, Man City, Chelsea, Arsenal, Liverpool, and Tottenham represented a group of teams that was higher in both popularity and team-strength than the rest, then unsurprisingly 10 out of the 11 matches with reduced profit margins reported by bwin include at least one such team; whereas only 2 out of the 6 biased matches reported for Ladbrokes do so. For bwin, this may suggest that it is likely the diminished profit margin introduced to those particular 11 matches was an attempt to attract more bettors due to the popularity associated with those matches. On the other hand, we have no strong evidence or strong rational assumptions to explain the behaviour of Ladbrokes. Regardless, it still comes to no surprise why none of those 6 particular matches featured a top-4[5] team; which may explain an identical activity to that of bwin (for less popular matches).

5.3.3. Arbitrage opportunities

Arbitrage possibilities depend on two factors: a) the divergence in outcome probabilities generated from normalised odds and b) the introduced margin over (a). If a set of HDA probabilities is found (for a single match instance) whereby the sum of probabilities within that set is <1, then a profit for the bettor can be guaranteed if the bets are placed such so that the return is identical whatever the outcome. For example, if we find that the best (lowest) probabilities for the bettor for a specific match instance, over a number of bookmaking firms, are $p(H) = 0.45$, $p(D) = 0.29$ and $p(A) = 0.25$, the sum of probabilities is just $sum = 0.99$; corresponding to the respective decimal odds of 2.2222, 3.4482 and 4. For this scenario we can guarantee a profit of $\left(\frac{1}{99}\right)$ 1.0101%. If we want to invest $bet = £100$, then the bet has to be distributed on the three outcomes as follows: $£45.4545$ on outcome H, $£29.2929$ on outcome D and $£25.2525$ on outcome A, using the following equation: $\frac{\left(\frac{bet}{sum}\right)}{odds}$ for each case of HDA (i.e. $odds$ equals the odds of H when calculating the bet to be placed on outcome H).

In previous sections, we have demonstrated that deviations in normalised probabilities are indifferent between different bookmakers, but significant in introduced profit margins. Further, given that the profit margins

[5] The top four teams in the English Premier League qualify for the UEFA Champions League. As noted in Section 5.2.3 Up until season 2010/11, those four places were consistently dominated by Manchester United, Chelsea, Arsenal and Liverpool.

have generally been dramatically reduced over the past few years (as illustrated in Section 5.3), and with no strong evidence of decrease in the divergence of those margins, it is certain that we are now in a better position to exploit positive arbitrage returns.

Academic evidence that demonstrate arbitrage opportunities date back to the 1980s, where Pope and Peel (1989) reported many such cases by considering the odds offered by four bookmakers, on a pre-tax basis, from 1980 to 1982. However, more recent studies (Dixon & Pope, 2004; Forest et al., 2005) found no such opportunities in modern online betting and concluded that there have been far less divergences in odds in recent years than in earlier periods. In particular, Forest et al. (2005) performed similar tests for the EPL seasons 1998 to 2003 using information from five bookmakers with introduced profit margins within the range of 10% and 12%. They showed that the minimum possible profit margin over the five seasons was averaging close to 6.6% and as expected, they reported no cases of positive arbitrage returns during that period.

For our tests, we have considered the EPL odds published by the seven bookmakers reported in Table 5.4 earlier for seasons 2004 to 2011. The combined minimum margin is reported in Table 5.5 along with the summary statistics for each of the 7 successive EPL seasons. As expected, the combined minimum margin steadily decreased over the period. This was enough to allow for a limited number (five) of arbitrage opportunities, as Tables 5.5 and 5.6 reveal. Table 5.6 reports on the five particular matches, the best combined odds along with the respective probabilities.

Table 5.5. Minimum combined profit margins achieved and arbitrage instances found between seven bookmakers, along with summary statistics for each EPL season considered.

EPL Season	Mean	Median	Standard Deviation	Arbitrage Instances	Mean profit
2004/05	0.0512	0.0530	0.0154	0	0.00%
2005/06	0.0499	0.0505	0.0130	0	0.00%
2006/07	0.0552	0.0572	0.0130	0	0.00%
2007/08	0.0441	0.0469	0.0150	2	0.12%
2008/09	0.0366	0.0366	0.0105	1	0.16%
2009/10	0.0364	0.0366	0.0107	0	0.00%
2010/11	0.0321	0.0330	0.0111	2	0.39%

Table 5.6. Details of the arbitrage instances found as described in Table 5.5.

Match	Date	Home Team	Away Team	Best combined odds						Sum of probab.
				H	D	A	$p(H)$	$p(D)$	$p(A)$	
1	29/09/2007	Chelsea	Fulham	1.40	5	12	0.7142	0.2000	0.0833	0.9976
2	25/11/2007	West Ham	Tottenham	2.90	3.60	2.65	0.3448	0.2777	0.3773	0.9999
3	19/10/2008	Hull	West Ham	2.62	3.40	3.10	0.3816	0.2941	0.3225	0.9983
4	27/11/2010	Stoke	Man City	3.75	3.40	2.30	0.2667	0.2941	0.4348	0.9956
5	26/02/2011	Wigan	Man Utd	7.20	4.30	1.60	0.1389	0.2326	0.6250	0.9964

5.4 FIXED-ODDS: Are they really fixed?

There is an assumption that football odds, which are typically first published one week before the match is played, remain fixed until the match starts. Indeed, all of the previous studies that have considered this issue have assumed or concluded a fixed-odds betting market. In particular, :

- (Pope & Peel, 1989; Forrest & Simmons, 2002; Goddard & Asimakopoulos, 2004; Forrest et al., 2005) claimed that the odds remain unaltered several days before the match even if new information is received;

- (Forrest et al., 2005) claimed that, although bookmakers retain the right to change the odds before the start of the match, they rarely do so;

- (Levitt, 2004) claimed that in sports betting generally adjustment of the odds are not only infrequent but also small when they occur.

In fact, contrary to the above, this section shows that a) adjustments in published odds do happen, b) they are frequent and c) they can be significant.

We provide an analysis on the adjustments of published odds observed by two bookmakers, Sportingbet and bwin. We have been monitoring the odds provided by each of the two bookmakers on a daily basis from 07/11/2010 to 09/05/2011, during which period 200 such cases have been recorded for the EPL matches. Apart from proving that such adjustments exist, the aim was also to understand a) the frequency of the adjustments, b) the significance of an adjustment, and c) the rationality behind the adjustments. In explaining (c), we have also attempted to record the potential causes behind each observed adjustment. Appendix A.1 provides this information. In summary, the rational causal factors do not really explain the changes so it is most likely that the volume of bets were the cause. Details regarding percentage shifts from initial to final central tendency distributions can be found in Appendix A.2 for all of the 200 occurrences. Table 5.7 presents six cases in which final published odds appear to have been dramatically altered from the odds that had been initially published, and Table 5.8 illustrates such an actual day-by-day adjustment example.

Table 5.7. Evidence of notable adjustments in published odds[6].

Date	Bookmaker	Home Team	Away Team	Initial published odds			Final published odds		
				H	D	A	H	D	A
22/11/2009	bwin	Bolton	Blackburn	1.75	3.45	4.50	2.20	3.25	3.10
16/01/2010	Sportingbet	Stoke	Liverpool	5.75	3.60	1.53	3.40	3.10	2.05
16/01/2010	bwin	Stoke	Liverpool	6.75	3.75	1.50	3.70	3.20	2.00
30/01/2010	Sportingbet	West Ham	Blackburn	2.10	3.20	3.20	2.80	3.20	2.30
16/02/2010	Sportingbet	Stoke	Man City	3.40	3.25	2.00	2.90	3.20	3.25
02/05/2010	Bwin	Liverpool	Chelsea	2.80	3.60	2.25	4.10	3.75	1.75

Further, Table 5.9 summarises the occurrence of adjustments per team/league position, and per predetermined intervals of group positions. A quick look reveals the tendency of bookmakers in providing more frequent adjustments for top teams than they do for bottom league teams. The difference in adjustments between upper and lower table appears to be significant[7] at 95% confidence interval. Assuming that teams at highest positions tend to be more popular and that the volume of bets on such matches is higher than the average, then this result agrees with the well known assumption of having bookmakers taking positions against bettors for maximising profit. Ultimately, only data from the volume of bets could confirm this, but unfortunately, bookmakers do not make such data publicly available.

[6] In 4 out of the 5 distinct matches presented only one bookmaker appears to have performed dramatic adjustments in published odds. This does not imply that the final published odds were dissimilar between the two bookmakers. The initial dates of odds publishing differs between bookmakers and, therefore, dramatic changes are most likely to occur to bookmakers who publish their odds very early.

[7] A two-tailed t-test was performed on two datasets with 200 Boolean indications. For instance, if the first dataset represents the upper table teams, return 1 (TRUE) at instance n if such a team is present at instance n, otherwise 0 (FALSE).

Table 5.8. An actual day-by-day adjustment example of bwin odds.

Match	Date of adjustment	Published odds			Normalised prob. distribution		
		H	D	A	H	D	A
Stoke Vs. Liverpool	11/01/2010	6.50	3.75	1.52	14.27	24.73	61.01
	14/01/2010	4.20	3.30	1.85	22.01	28.02	49.97
	15/01/2010	3.40	3.20	2.10	27.16	28.86	43.98
	16/01/2010	3.70	3.20	2.00	24.96	28.86	46.18
	10/01/2010	6.75	3.75	1.50	13.70	24.66	61.64

Table 5.9. Adjustment of the odds observed per league position.

League position	Team	Adjustments	Adjustments per league interval	Adjustments per league interval
1	Chelsea	17	54.75% (upper table)	27.50% (pos. 1-5)
2	Man United	20		
3	Arsenal	28		
4	Tottenham	20		
5	Man City	25		
6	Aston Villa	23		27.25% (pos. 6-10)
7	Liverpool	26		
8	Everton	24		
9	Birmingham	18		
10	Blackburn	18		
11	Stoke	20	42.25% (lower table)	22.50% (pos. 11-15)
12	Fulham	18		
13	Sunderland	17		
14	Bolton	24		
15	Wolves	11		
16	Wigan	20		22.75% (pos. 16-20)
17	West Ham	18		
18	Burnley	17		
19	Hull	15		
20	Portsmouth	21		

5.5 Discussion and implications

An attempt was made to assess the degree of inefficiency of the football betting market. We have considered the odds of 7 different well known bookmakers provided for 9 football leagues from 4 different countries, for period 2000 to 2011. Our findings are summarised as follows:

5.5.1. Accuracy

The results lead us to agree with (Forrest et al., 2005) who claimed that the variation in the predictability of match results from season to season is much larger than the variation in forecasting performance between bookmakers for the same season. However, whereas (Forrest et al., 2005) concluded that the notion of bookmakers providing more accurate odds over time is probably correct, our results show that this notion is false since over a period of 11 years no evidence of forecast improvement have been observed.

Also, we have showed that bookmakers normally perform worse for lower divisions than they do for top divisions within the same country. However, this cannot be explained by either the popularity or the strength of a league. This is because our results show that bookmakers' performance was significantly better for the top Greek division than it was for the top English and Spanish divisions, and this is consistent for almost a decade. Since top divisions in England and Spain are undoubtedly superior in both quality and popularity than that of Greece, this behaviour can only be explained by the predictability of the league. This assumption is also backed up by further evidence in Table 5.2 (Section 5.2.3) which shows that this phenomenon occurs to the same league for different seasons.

Further, Forrest et al. (2005) found that William Hill was the best performing bookmaker, out of a total of five, for period 1998 to 2003. Our results show that the best performing bookmaker was BET365, for the EPL period 2007 to 2011. In our study William Hill is ranked 3rd. However, none of the two bookmakers ranked above William Hill in this study have been considered in (Forrest et al., 2005), and Ladbrokes is ranked worse than William Hill in both studies.

5.5.2. Favourite-longshot bias

Contrary to the claims[8] of (Dixon & Pope, 2004), we have found evidence of the well known phenomenon called the favourite-longshot bias. We have demonstrated that bookmakers offer less than fair odds on away wins than on home wins, and this observation is consistent over a 7-year period. Our conclusions agree with those in (Cain et al., 2000; Forrest et al., 2005; Graham & Stott, 2008). We have no evidence to explain this irrationality, but a good possibility is that bookmakers take dynamic positions against the presumed tendency of the bettors to underbet on favourites and to overbet on

[8] Even though Dixon and Pope (2004) appear to demonstrate identical results to those which follow the so called favourite-longshot bias, they made claims of a favourite bias (or reverse long-shot bias). In particular, they claim that "*Fig. 7 suggests that the fixed odds contain a favourite bias, i.e. a reverse long-shot bias. The odds on low probability (long-shot) outcomes are too generous, and those on high probability outcomes are too short. This conclusion is reinforced by remembering that the benchmark returns from betting on all homes or all always are* −8% *and* −14%*, respectively*" (Dixon & Pope, 2004).

risky outcomes, as also suggested in (Rossett, 1971; Snyder, 1978; Ali M. M., 1979; Asch et al., 1984; Levitt, 2004; Graham & Stott, 2008).

5.5.3. Profit margin

Our findings confirm the assumptions made in (Rue & Salvesen, 2000) who suggested that it is natural for the bookmakers to provide better odds for the Premier League than for the lower divisions, as the majority of the bettors bet on the Premier League. Further, our results show that introduced profit margins appear to diminish after each successive season, and are consistent with (Hvattum & Arntzen, 2010) who showed that the competitiveness of the football betting market has increased during period 2000 to 2008.

However, we have showed that significantly different profit margins can be introduced per bookmaker (which is also verified by arbitrage opportunities below), implying that the odds of a single bookmaker cannot represent the overall market. This contradicts the suggestions in (Forrest & Simmons, 2002; Forrest et al., 2005).

5.5.4. Arbitrage opportunities

By combining bets on different outcomes of a match with different bookmakers a bettor can considerably reduce. his exposure to risk. Moreover, contrary to the claims of recent academic papers, we have demonstrated that there continue to be a small number of arbitrage opportunities which allow risk-free profits at zero cost. We have considered seven bookmakers with mean profit margins ranging from approximately 5.5% to 12.5% over seven EPL seasons, and a combined minimum close to around 3%, which was predictably observed during the most recent season. This exposed five arbitrage instances to guarantee profits up to approximately 0.5% per instance.

What makes this finding particularly important is that we have only used a tiny fraction of the available information to identify such opportunities. Specifically:

1. We have only considered seven bookmakers, whereas even in 2006 there were already 467 online bookmakers (Top 100 bookmakers, 2006) reports on;

2. We only considered one league (the EPL), whereas there are several hundred on which most bookmakers lay odds;

3. We have only considered one type of bet; outcomes $\{p(H), p(D), p(A)\}$;

4. We have not included Asian markets where profit margins go as low as 1% (Graham & Stott, 2008);

5. We have not considered betting exchange markets such as Betfair (betfair, 2000) whereby one bettor bets against another for a mutually agreed price;

6. We have not considered taking into advantage the large number of different bonuses offered per bookmaker;

7. We have not made use of the irrational frequent adjustments of published odds (see Sections 5.4 and 5.5).

Clearly, by taking account of all of the above, there must be many more arbitrage opportunities, with significantly increased guaranteed profit, than those we have identified. Such opportunities will increase in the future as profit margins get reduced even further due to competitiveness (see Section 5.3). Finally, there is the emergence of software to make it easier to spot arbitrage opportunities; for example, websites such as www.oddschecker.com[9] make the whole process much easier and there are also evolving systems that perform automated internet analysis in real time to spot arbitrage opportunities.

All of this highlights bookmakers' exposure to substantial risks. Assuming that the betting markets allow such inefficiencies for commercial purposes, it is evident that at some point bookmakers will be forced to eliminate them if they are to retain maximum profitability.

5.5.5. Adjustment of published odds

In contrast to many previous studies, we have demonstrated that adjustments in published odds exist, they are frequent, and they can be significant. Further, we have also demonstrated that a) such adjustments can be irrational since bookmakers appear to introduce conflicting adjustments for identical events on the same day, and b) the preference of bookmakers in

[9] A website that gives an overall view of the market and informs visitors about the best available odds by considering a large number of various online bookmakers.

providing more frequent adjustments for matches with higher popularity; which may imply that that they take positions in maximising profit due to a possibly increased volume of bets received.

5.5.6. Objective, subjective and extraneous information

In (Webby & O'Connor, 1996; Forrest et al., 2005; Graham & Stott, 2008) the authors suggest that bookmakers are privy to, and make effective use of, information not captured by their statistical models (both subjective and objective). This suggestion may be the result of their statistical models failing to perform as well as bookmakers' odds did, and/or after noticing that the forecast accuracy of their model is improved by passing bookmakers' odds as one of the model parameters. Further, in Chapter 6 we provide strong evidence to back this suggestion by demonstrating how our football forecasts were revised from being statistically different to being statistically indifferent against normalised published odds, after incorporating relevant expert knowledge and statistical analysis for generating football forecasts before the matches are played.

Results from this study (mainly Section 5.4) demonstrate that bookmakers make effective use of both objective and subjective information, but they also appear to introduce extraneous information which cannot be explained. However, the room for improvement in the manipulation of such information is evident by the fact that even bookmakers fail to adjust quickly enough from new evidence introduced in unpredictable leagues (Sections 5.2.2 and 5.2.3). This already suggests the limitation of data-only approaches to prediction.

5.6 Concluding remarks and future work

We conclude that the football betting market is inefficient, particularly in the presence of arbitrage opportunities, regular predetermined biases in published odds, and conflicting bookmaking adjustments in published odds. However, we consider this inefficiency as the outcome of commercial objectives rather than lack of ability. In particular, the gambling market appears to allow exposure and losses against the very best of bettors and in return increases profits against the residual, more causal bettors. Given that not many of the bettors are actual professionals, it is not unreasonable to assume that bookmakers can still afford the luxury of such an exposure. It seems that bookmakers continue to generate huge profits because the typical bettor:

- plays for pleasure;
- is lazy and wants to stick with one bookmaker;
- is greedy (tends to immediately rebet winnings).

But above all a typical bettor is ignorant of the relevant risks governing his various betting scenarios. A bettor with a certain level of ability, who can also treat football betting as no different to stock trading, should be in a position to beat this market at a consistent rate.

Further, if the assumption of having bookmakers taking positions against bettors for maximising profit is correct, then bookmakers' odds are prices published with the intention of maximising profit; implying that such odds are not to be interpreted as truthful forecasts for assessment. For instance, the well known tendency of the favourite-longshot bias should be taken into consideration prior to deriving conclusions since it is almost certain that a betting strategy which supports bets on favourite outcomes will generate higher returns than another which does not. However, this is completely ignored (primarily) in all of the previous studies that introduce football forecast models.

The results here suggest many possible directions for future work. What appears to be missing from the academic literature is how bettors can take advantage of the various bonuses (e.g. deposit bonus) offered by many bookmakers. Further, almost all of the past studies have only focused on $\{p(H), p(D), p(A)\}$ odds for deriving conclusions, primarily due to availability limitations. It would be very interesting to investigate how betting markets behave for bets other than the standard football outcomes (e.g. players, goal-lines, cards, correct scores, tournament outrights etc.). Finally, and probably most important, could be an investigation in how markets behave during live betting. Live betting has emerged along with online betting it is has now become exceptionally popular, with bookmakers reporting that live betting accounts for the majority of the betting stakes, or approximately 75% of the total as reported in (bwin Group, 2010) which in turn represents a growth of approximately 7.1% from the previous year.

CHAPTER 6

pi-football Model $v1.32$: A Bayesian network model for forecasting Association Football match outcomes

The novel material introduced in this chapter comes from our publication (Constantinou et al., 2012a), and presents a BN model for forecasting football match outcomes in which subjective variables represent the factors that are important for predictions but which historical data fails to capture. This was the first publication to demonstrate profitability that was consistent against all of (the available) bookmakers' odds over a large period of time.

6.1 Introduction

In this chapter we present a Bayesian network model for forecasting the outcomes of football matches in the distribution form of $\{p(H), p(D), p(A)\}$; corresponding to home win, draw and away win, whereby the subjective variables represent the factors that are important for prediction but which historical data fails to capture. The model (pi-football) was used to generate forecasts about the outcomes of the English Premier League (EPL) matches during season 2010/11 (but is easily extended to any football league). Forecasts were published online prior to the start of each match. We believe this study is important for the following reasons:

1. using an appropriate measure of forecast accuracy, the subjective information improved the model such that posterior forecasts were on par with bookmakers' performance;

2. using a standard profitability measure with discrepancy levels at $\geq 5\%$ the model is profitable under maximum, mean and common bookmakers' odds, even by allowing for the bookmakers' introduced profit margin;

3. the model priors are dependent on statistics derived from predetermined scales of team-strength, rather than statistics derived from a particular team (hence enabling us to maximise historical data);

4. the model enables us to revise forecasts from objective data, by incorporating subjective information for important factors that are not captured in the historical data;

5. the significance of recent information (objective or subjective) is weighted using degrees of uncertainty resulting in a non-symmetric Bayesian parameter learning procedure;

6. forecasts were published online at www.pi-football.com before the start of each match (pi-football, 2010);

7. although the model has so far been applied for one league (the English Premier League) it is easily applicable to any other football league.

Hence, compared with other published football forecast models, pi-football not only appears to be exceptionally accurate, but it can also be used to 'beat the bookies'. Even though this model is replaced by a better, simpler model in the next chapter we nevertheless present it here in order to demonstrate how 'lessons learned' from this first model lead to the superior model presented in Chapter 7.

This Chapter is organised as follows: Section 6.2 describes the historical data and method used to inform the model priors, Section 6.3 describes the Bayesian network model, Section 6.4 describes the assessment methods along with results and discussion, and Section 6.5 provides our concluding remarks and future work.

6.2 Data

The basic data used to inform the priors for the model were the results (home, draw or away) of all English Premier League (EPL) matches from season 1993/94 to 2009/10 inclusive (a total of 6244 occurrences). This information is available online at (Football-Data, 2012). The forecasts generated by the model were for season 2010/11, a total of 380 EPL matches.

In contrast to previous approaches we use the historical data to generate prior forecasts that are 'anonymous' by using predetermined levels of team-strength, rather than distinct team-names. We achieve this by replacing each team-name in each match in the database with a ranked number that represents the strength of that particular team for a particular season. The team-strength number is derived from the total number of points[1] that the particular team achieved during that particular season as shown in Table 6.1.

Table 6.1. Predetermined levels of team strength.

Total points	>84	80-84	75-79	70-74	65-69	...(intervals of 5 points)	30-34	25-29	<25
Strength	1	2	3	4	5	...	12	13	14

This implies that the same team may receive different ranks for different seasons and that different teams may receive identical ranks within the same season.

[1] In EPL a total of 20 football teams participate and thus, a team can accumulate a minimum of 0 and a maximum of 114 points.

For example, the Manchester City at home to Aston Villa match in season 2006-07 is classified as a ranked 10 versus a ranked 8 team (because in that season Manchester City totalled 42 points and Aston Villa 50 points), whereas in season 2009-10 the Manchester City at home to Aston Villa match is classified as a ranked 5 versus a ranked 6 team (because in that season Manchester City totalled 67 points and Aston Villa 64 points).

The granularity (of 14 levels of team strength) has been chosen to ensure that for any match combination (i.e. a team of strength x at home to a team of strength y) there are sufficient data points for a reasonably well informed prior for $\{p(H), p(D), p(A)\}$. This approach has a number of important advantages:

a) it enables us to make maximum use of limited data and be able to deal with the fact that every season the set of 20 teams changes (three are relegated and three new teams are promoted). For example, forecasts for teams for which there is little or no historical data (such as those recently promoted) are based on data for different teams but of similar strength;

b) historical observations do not have to be ignored or weighted since the challenge here is to estimate a team's current strength and *learn* how such a team performed in the past given the specified ground (home/away) and opponent's strength. For example, consider the prior for the Manchester City at home to Aston Villa match in season 2010-11. Because the historical performances of Manchester City and Aston Villa prior to season 2010-11 were in no way representative of their strength in season 2010-11, what matters is not the results of previous matches between Manchester City and Aston Villa (which would be sparse as well as irrelevant), but the results of all previous matches where a rank 4 team played at home to a rank 9 team;

c) historical observations do not necessarily require weekly updating. The database already consists of thousands of historical match observations, and adding a few more matches every week will not make a major difference (this can be done once a year);

d) historical observations from one league can be used to predict match results for teams in another league (as long as the introduced ranking is redefined to accommodate potential discrepancies in the number of teams participating within that league).

6.3 The model

The model, which we call 'pi-football' ($v1.32$), generates predictions for a particular match by considering four generic factors for both the home and away team, namely: 1) strength, 2) form, 3) psychology and 4) fatigue. The factors (1) and (2) are known to be particularly important when predicting football outcomes (Knorr-Held, 2000; Hvattum & Arntzen 2010; Leitner et al., 2010), factor (3) was selected arbitrarily as an additional 'test-case' factor, whereas factor (4) was selected on the basis of reduced player performance demonstrated within the sports science literature based on evidence of fatigue (Krustrup & Bangsbo 2001; Krustrup et al., 2003; Mohr et al., 2004; Castagna et al., 2006; Krustrup et al., 2006; Mujika et al., 2007; Mujika et al., 2008). There are model components corresponding to each of the four generic factors.

In this sections we describe each of the model components (with further details regarding the assumptions and the different scenarios available for each of the Bayesian network nodes provided in Appendix B.1), but first we provide a brief overview.

Component 1 provides an estimate of each team's current strength (based on recent data) expressed as a distribution. Using historical outcomes between such ranked teams we get a distribution for the predicted outcome as shown in Figure 6.1. Here we have a home team with mean strength 65-69 points (or rank 5) and an away team with mean strength 80-84 points (or rank 2). Component 1 is predominantly dependent on objective information for prediction and thus, we will refer to the resulting forecasts as 'objective forecasts'.

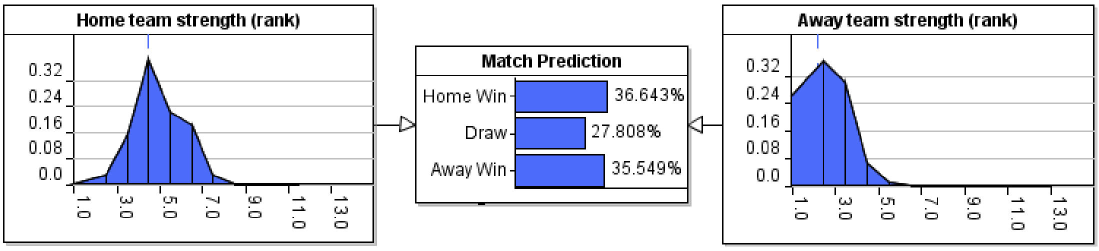


Figure 6.1. An example of an objective forecast generated at component 1.

Components 2, 3 and 4 are predominantly dependent on subjective information. They are used to revise the forecast from component 1. The outcome of each of the components is mutually summarised in a single value (considering both teams) which we describe as 'subjective proximity'. The subjective proximity is measured on a scale from 0 to 1. A value equal to 0.5 indicates no advantage to either of the teams; a value less than 0.5 indicates an advantage for the home team, while a value greater than 0.5 indicates an advantage for the away team. Since the forecast nodes are ranked in the sense of (Fenton et. al., 2007), the Bayesian Network software we have used

(Agena, 2012) automatically updates the forecast taking account of the subjective proximity as shows for different examples in Figure 6.2. Figure 6.3 illustrates how the four components are linked. We will refer to the revised (and final) forecasts as 'subjective forecasts'.

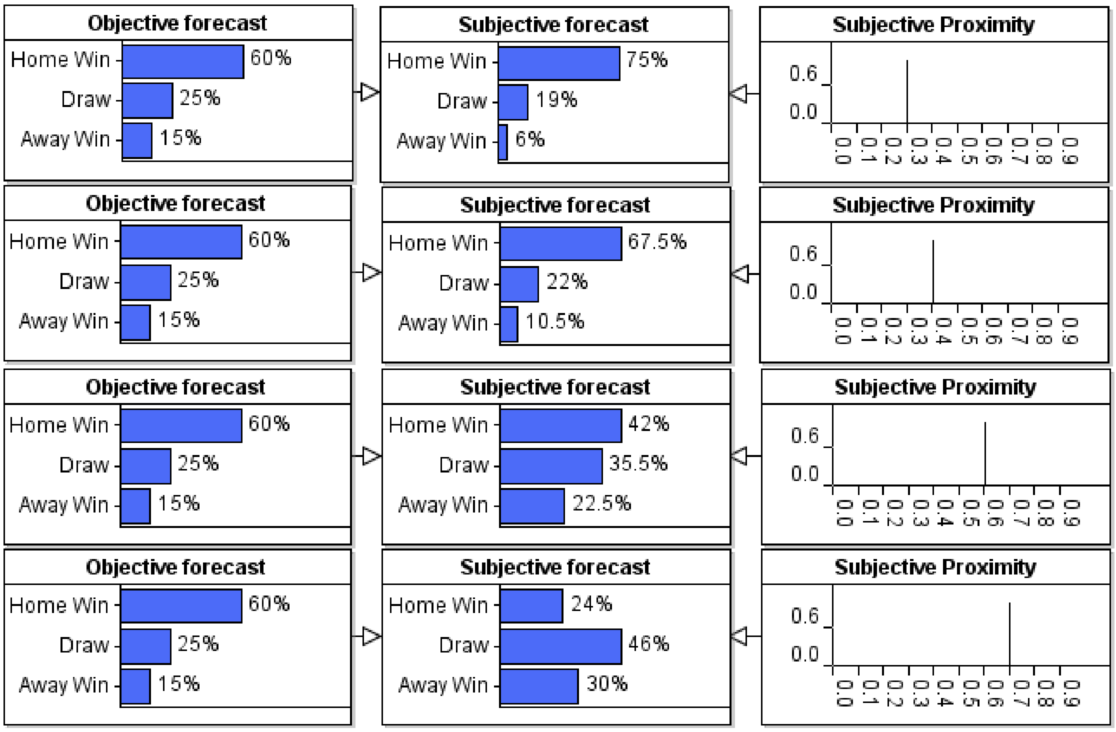


Figure 6.2. Forecast revision given different indications of subjective proximity.

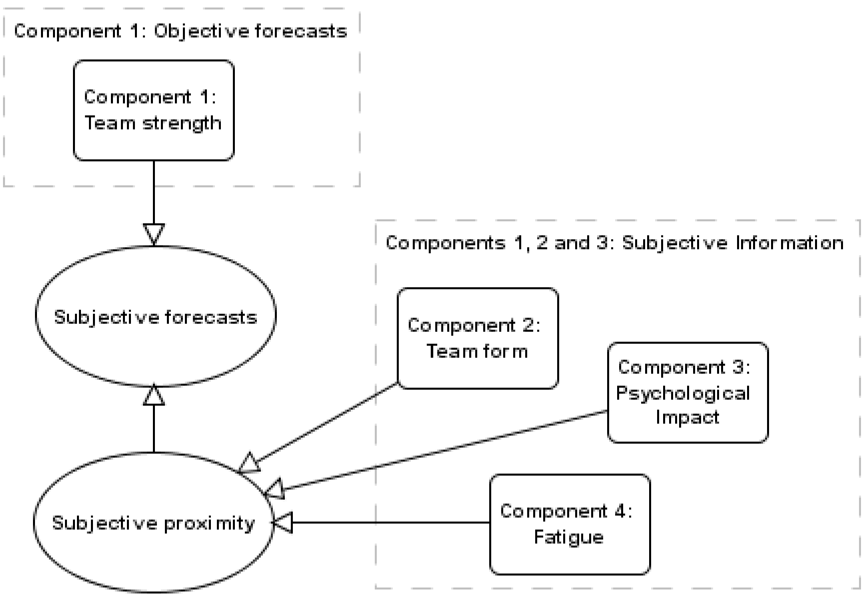


Figure 6.3. How components 1, 2, 3 and 4 are linked.

6.3.1. Component 1: Team strength

The Bayesian network corresponding to the team strength component is shown in Figure 6.4 and it can be explained in terms of the following information:

a) *Previous information*: represented by five parameters (nodes 2, 3, 4, 5, and 6), each of which holds the number of total points accumulated in each of the five previous seasons with degrees of uncertainty (higher uncertainty for older seasons);

b) *Current information*: represented by a single parameter (node 9) that holds an estimate about the strength of the team in total points, and which is measured according to the total points accumulated during the current season and the points expected from residual matches[2] with degrees of uncertainty (lower uncertainty for higher number of matches played);

c) *Subjective information (optional):* represented by a single parameter (node 7) that holds the expert's subjective belief about the strength of the team in total points with degrees of uncertainty (reflects the expert's confidence). This information is used in cases where important changes happen before the start of the current season that cannot be captured by the historical data. A good example is Manchester City at

[2] It is important to appreciate that the resulting parameter summarises a belief about the team's strength in points and not the points the team is expected to have by the end of the proceeding season.

the start of seasons 2009/10, 2010/11 and 2011/12, who dramatically improved their strength by spending $£160m$, $£77m$ and $£75m$ respectively signing some of the world's top players (Soccer Base, 2012).

The degree of uncertainty is modelled by exponential predetermined levels of variance in an attempt to achieve a limited memory process. This process produces a non-symmetric Bayesian parameter learning procedure. Accordingly,

a) *Previous information*: this indication receives increased rates of variance (and hence become less important) for each previous season, following the exponential growth illustrated in Figure 6.5a;

b) *Current information*: this indication receives decreased rates of variance (and hence become more important) after each subsequent gameweek[3], following the exponential decay illustrated in Figure 6.5b;

c) *Subjective Information:* this indication receives decreased or increased rates of variance according to the expert's confidence regarding his indication. The decreased/increased rates of variance follow those of the *previous information*[4] (Figure 6.5a).

[3] A complete EPL season consists of 38 gameweeks.

[4] For example, the degree of uncertainty when the expert's confidence is "Very Low" (fifth lowest out of five) is equal to the degree of uncertainty introduced for the points accumulated during the 5th preceding season.

Further information regarding the variables and available scenarios of this process is provided in Table B.1.1. An example with observations from the actual match between Man City and Man United dated 10th of November 2010 is illustrated in Appendix B.2.

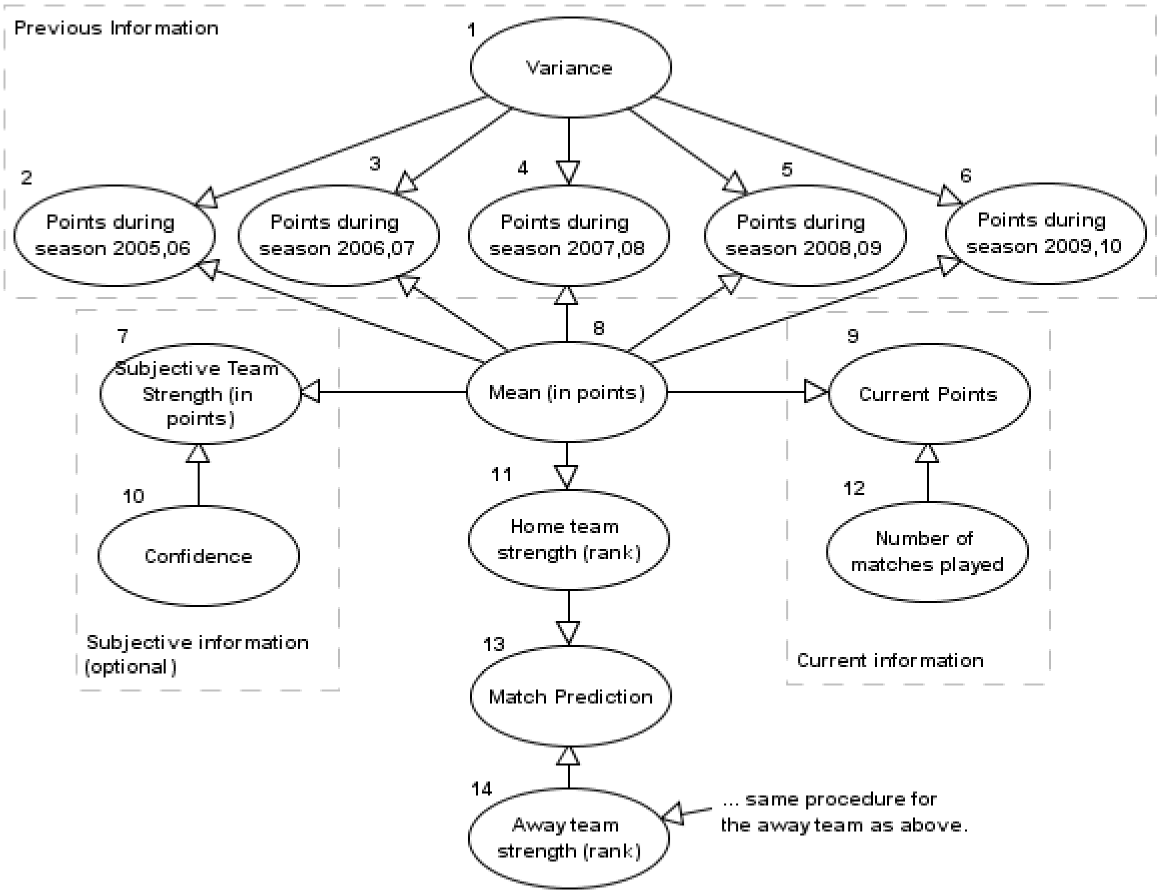
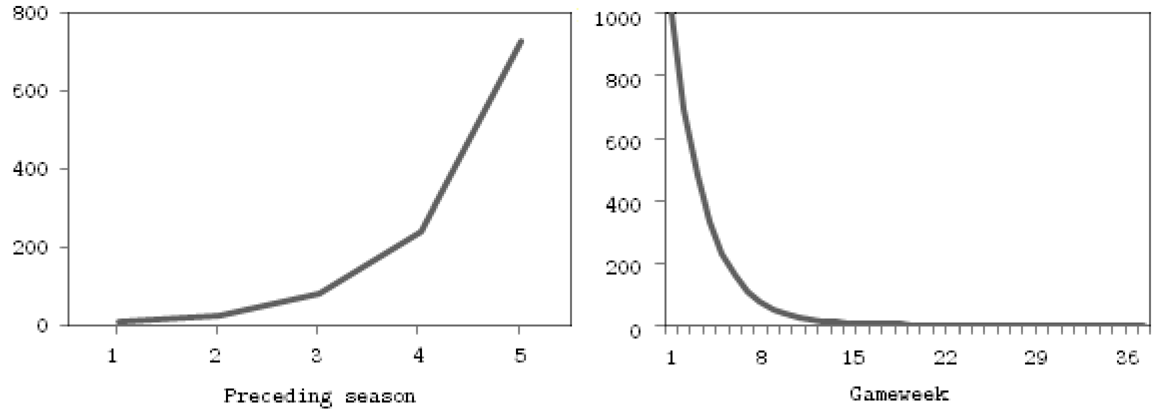


Figure 6.4. Component 1: Non-symmetric Bayesian parameter learning network for measuring the strength of the two teams and generating objective match predictions



(a) (b)

Figure 6.5. Limited memory process achieved by exponential growth/decay rates of uncertainty for the (a) previous seasons and (b) gameweeks played under the current season.

6.3.2. Component 2: Team form

This Bayesian network component is shown in Figure 6.6. The 'form' of a team (node 10 for the home team and 12 for the away team) indicates the particular team's recent performance against expectations, and it is measured by comparing the team's expected performance[5] against its observed performance during the five most recent gameweeks.

The form of a team is represented on a scale that goes from 0 to 1. When the value is close to 0.5 it suggests that the team is performing as expected; a higher value indicates that the team is performing better than expected. Further, if the particular team is playing at home, then the model will consider home form and away form with subjective weights $\left[\frac{2}{3}, \frac{1}{3}\right]$ respectively (nodes 5, 6, 7; the reverse applies for the away team). The form is revised according to subjective indications about the availability of certain

[5] Represented by what the model had initially forecasted.

players (nodes 1, 2, 3, 4)[6]. The expert constructed Bayesian network determines whether one team has an advantage over the other when comparing each other's form. Further information regarding the variables and available scenarios of this process is provided in Table B.1.2.

Figure 6.6 Component 2: Expert constructed Bayesian network for estimating potential advantages in form between the two teams.

[6] Form decreases if the team has new first-team injuries and increases when important players return back to action.

6.3.3. Component 3: Psychological impact

This Bayesian network component is shown in Figure 6.7. The psychology of a team is determined by subjective indications regarding motivation, team spirit, managerial issues and potential head-to-head biases. The Bayesian network estimates the difference in psychological impact between the two teams. This process is divided into two levels; where the information assessed during level 1 (node 6) is updated at level 2 (node 7). This implies that the total information of level 1 (nodes 1, 2) shares identical impact with that of level 2 (node 4). Further information regarding the variables and available scenarios of this process is provided in Table B.1.3.

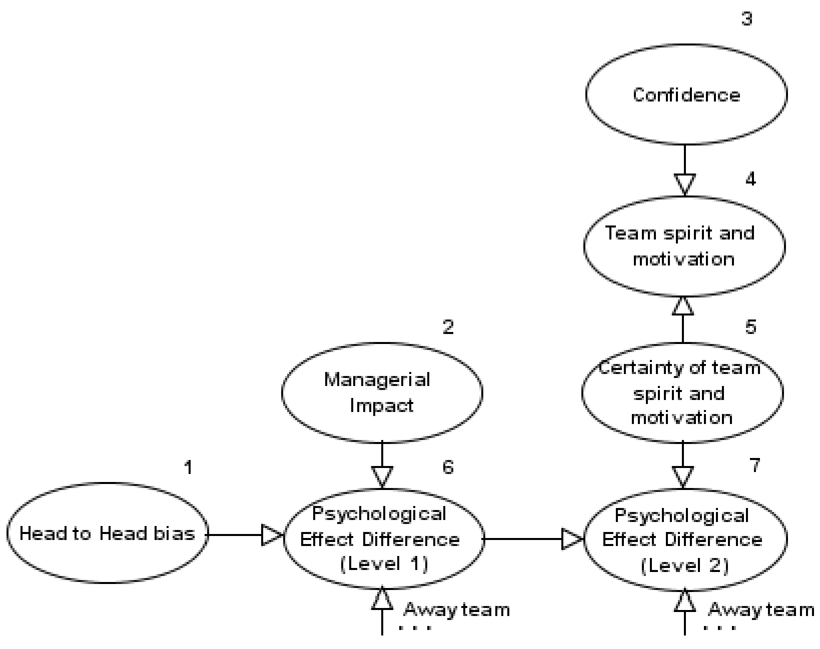


Figure 6.7. Component 3: Expert constructed Bayesian network for estimating potential advantages in psychological impact between the two teams.

6.3.4. Component 4: Fatigue

This Bayesian network component is shown in Figure 6.8. The fatigue of a team is determined by the toughness of the previous match, the number of days gap since that match, the number of first team players rested (if any), and the participation of first team players in national team matches (if any). The Bayesian network estimates the difference in the level of fatigue between the two teams. In particular, the resulting tiredness, which is determined according to the toughness of the previous match (node 5), is diminished according to a) the number of days gap since the last match (node 1), and b) the number of first-team players rested during that match[7] (node 2). Further, the indication of fatigue may increase up to 50% towards its maximum value depending on the level of participation of first team players in additional matches with their national team[8] (nodes 6, 7). If there is no national team participation the fatigue will receive no increase. Further information regarding the variables and available scenarios of this process is provided in Table B.1.4.

[7] Where (a) is defined to be twice as important to (b) when calculating 'Restness' (node 3).

[8] When football teams are given a break due to national matches, top level teams (e.g. Man United) might suffer greater levels of fatigue due to having many players who are first-team regulars with their national team.

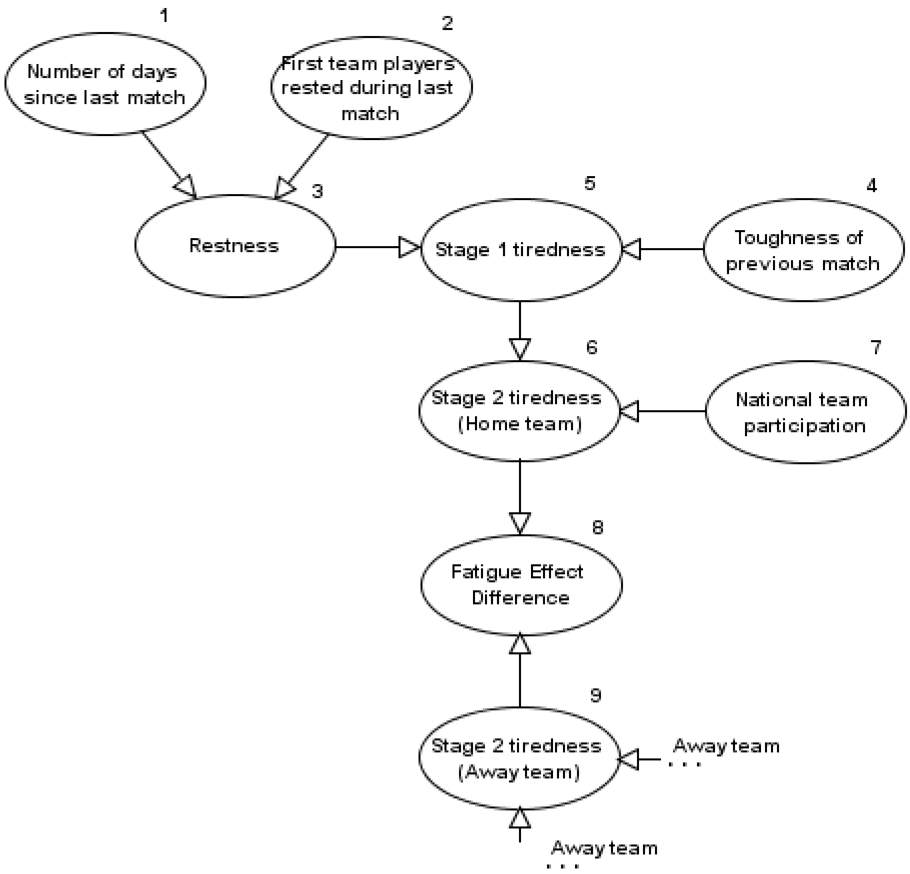


Figure 6.8. Component 4: Expert constructed Bayesian network for estimating potential advantages in fatigue between the two teams.

6.4 Results and discussion

There are various ways in which the quality of a forecast model can be assessed. In particular, we can consider *accuracy* (how close the forecasts are to actual results) and *profitability* (how useful the forecasts are when used as the basis of a betting strategy). Researchers have already concluded that there is only a weak relationship between commonly used measures of accuracy and profitability (Leitch & Tanner, Economic Forecast Evaluation: Profits Versus The Conventional Error Measures, 1991) and that a combination of the two might be best (Wing et. al., 2007). Hence we use assessments of both accuracy

(Section 6.4.1) and profitability (Section 6.4.2) in order to get a more informative picture about the performance of pi-football, whereas Section 6.4.3 provides an analysis of impact of the subjective components of the model based on the two measures.

6.4.1. Accuracy Measurement

For assessing the accuracy of the forecasts we use of the Rank Probability Score (RPS). We explained why it was the most rational scoring rule of those that have been proposed and used for football outcomes in Chapter 4.

To determine the accuracy of our model we compute the RPS for the following three forecasts:

a) the objective forecasts generated at component 1; we will refer to these forecasts as f_O;
b) the subjective (revised) forecasts after considering components 2, 3 and 4; we will refer to these forecasts as f_S;
c) the respective normalised[9] bookmakers' forecasts; we will refer to these forecasts as f_B.

Other studies have concluded that the normalised odds of one bookmaker are representative of any other bookmaker (Dixon & Pope, 2004; Forrest et al., 2005); we also demonstrate this in Chapter 5. However, instead of selecting a

[9] The bookmakers' odds are normalised such so that the sum of probabilities over the possible events is equal to 1 (the introduced profit margin is eliminated). For more information see Chapter 5.3.

single bookmaker we make use of the mean[10] bookmakers' odds as provided by (Football-Data, 2012). Figure B.3 demonstrates the RPS generated per forecast under the three datasets.

Figure 6.9 presents the cumulative RPS difference for a) $f_B - f_O$, b) $f_B - f_S$, and c) $f_O - f_S$. Since a higher RPS value indicates a higher error a cumulative difference for $A - B$ below 0 indicates that A is more accurate than B. Accordingly, the graphs suggest that the accuracy of pi-football improves after considering subjective information. However, the bookmakers appear to have a higher overall accuracy even after the forecasts are revised. We performed 2-tailed paired t-tests to determine the importance of the above discrepancies. The null hypothesis is that the two datasets are represented by similar forecasts. The results are:

a) the dependence between dataset f_O and dataset f_B is statistically significant at 99% confidence interval with a p-value of 0.0023; therefore, the null hypothesis is rejected;
b) the dependence between dataset f_S and dataset f_B is statistically significant at 99% (not even at 90%) confidence interval with a p-value of 0.1319; therefore, the null hypothesis fails to be rejected.

We conclude that the accuracy of objective forecasts was significantly inferior to bookmakers' forecasts, and that subjective information improved the forecasts such that they were on par with bookmakers' performance. This also suggests that the bookmakers, as in the pi-football model, make use of information that is not captured by the standard statistical football data

[10] The mean odds are measured by considering a minimum of 28 and a maximum of 40 different bookmakers per match instance (Football-Data).

available to the public. Further, appendix B.4 provides evidence of significant improvements in f_O by incorporating subjective information. Table B.4.1. presents match instances in which f_O and f_S generate the highest RPS discrepancies, along with indications whether f_S lead to a more accurate forecast.

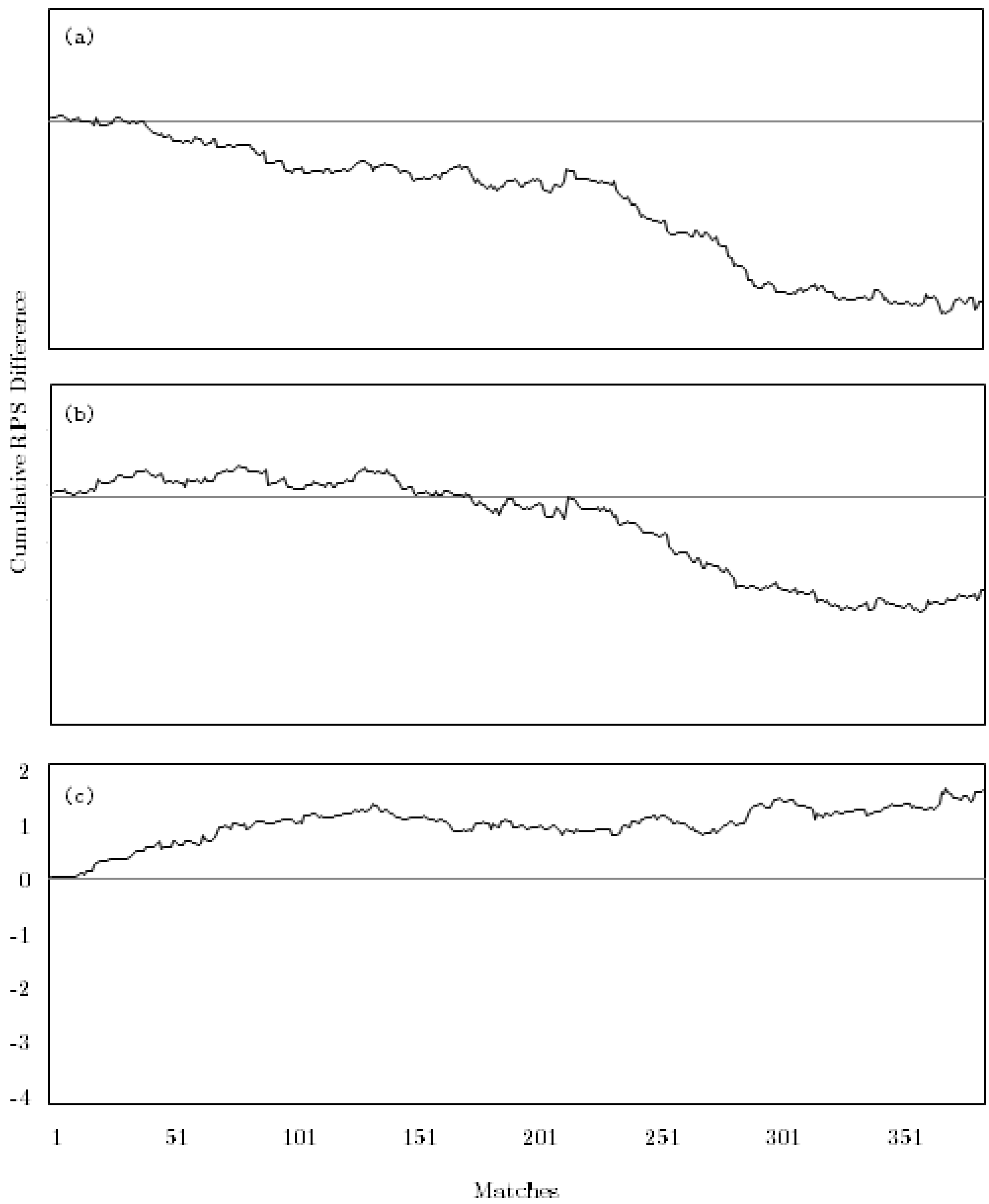


Figure 6.9. Cumulative RPS difference when (a) $f_B - f_O$, (b) $f_B - f_S$, (c) $f_O - f_S$. Since a higher RPS value indicates a higher error a cumulative difference for $A - B$ below 0 indicates that A is more accurate than B.

6.4.2. Profitability Measurement

For assessing the profitability of the forecasts we perform a simple betting simulation which satisfies the following standard betting rule: *for each match instance, place a* £1 *bet on the outcome with the highest discrepancy, of which the pi-football model predicts with higher probability, if and only if the discrepancy is greater or equal to* 5%.

This assessment, of course, depends on the availability of an appropriate bookmaker's odds[11]. In contrast to previous papers (Forrest & Simmons, 2002; Forrest et al., 2005), our work in Chapter 5.3.2 shows that the published odds of a single bookmaker are not representative of the overall market. Unlike the case of accuracy (Section 6.4.1) where published odds are normalised and hence the profit margin is eliminated, for profitability we have to consider the published odds (such odds are not normalised and are considered with their profit margins), hence the odds of one bookmaker can be significantly different to another. Accordingly, in determining pi-football's profitability we consider the following three different sets of bookmakers' odds[12]:

a) *the maximum* (best available for the bettor) bookmakers' odds which we are going to refer to as f_{maxB}. This dataset is used to estimate how

[11] See also the following studies on the football gambling market: (Pope & Peel, 1989; Dixon & Coles, 1997; Kuypers, 2000; Rue & Salvesen, 2000; Forrest & Simmons, 2001; Dixon & Pope, 2004; Goddard & Asimakopoulos, 2004; Forrest & Simmons, 2008; Graham & Stott, 2008).

[12] The bookmakers' odds are also provided by (Football-Data).

an informed bettor, who knows how to pick the best odds by comparing the different bookmakers' odds, could have performed;

b) *the mean* (average) bookmakers' odds which we are going to refer to as f_{meanB}. This dataset is used to estimate how an ignorant bettor could have performed, assuming he selects a bookmaker at random;

the *most common* bookmakers' odds which we are going to refer to as f_{WH}. This dataset is used to estimate how the common UK bettor could have performed. For this, we consider the odds provided by the leading UK bookmaker William Hill, who represents the 25% of the total market throughout the UK and Ireland (William Hill PLC, 2012).

Figure 6.10 demonstrates the cumulative profit/loss generated against a) f_{maxB}, b) f_{meanB} and c) f_{WH} after each subsequent match, assuming a £1 stake when the betting condition is met. The model generates a profit under all of the three scenarios and the simulation almost never leads into a negative cumulative loss even allowing for the in-built bookmakers' profit margin[13]. Figure 6.11 illustrates the *Risk of Ruin* for up to a bankroll 100 times the value of a single bet. A bankroll of approximately £55 (or 55 times the value of a single bet) and approximately £45 is required to ensure that the probability to lose the specified bankroll under infinite betting is $\leq 5\%$ for f_{maxB} and f_{WH} respectively. In the case of f_{meanB} the profit rate is not high enough to ensure a risk of ruin $\leq 5\%$ with a bankroll up to 100 times the

[13] We have also performed the identical betting simulation given f_O. Figure B.5 demonstrates how the betting simulation results in losses of -13.98% against f_{maxB}, -19.92% against f_{meanB} and -12.84% against f_{WH}. This confirms the accuracy measurement results; that is, the significant improvements in f_O (which formulate f_S) by incorporating subjective information.

value of a single bet. Table 6.2 summarises the statistics of the betting simulation for all of the three scenarios.

Overall, pi-football won approximately 35% of the bets simulated under all of the three scenarios, with the mean odds of winning bets at approximately 3.00. This suggests that the model was able to generate profit via longshot bets; what makes this especially interesting is that longshots are proven to be biased against the bettors (Cain et al., 2000, Forrest & Simmons, 2001; 2002; Forrest et al., 2005; Graham & Stott, 2008), as we also demonstrate this in Chapter 5. This implies that the model would have generated even higher profits if the betting market was to provide unbiased odds. Additionally, profits are most likely to have been even higher under scenarios (b) and (c) if we were to eliminate the respective built-in profit margins of 6.09% and 6.50%.

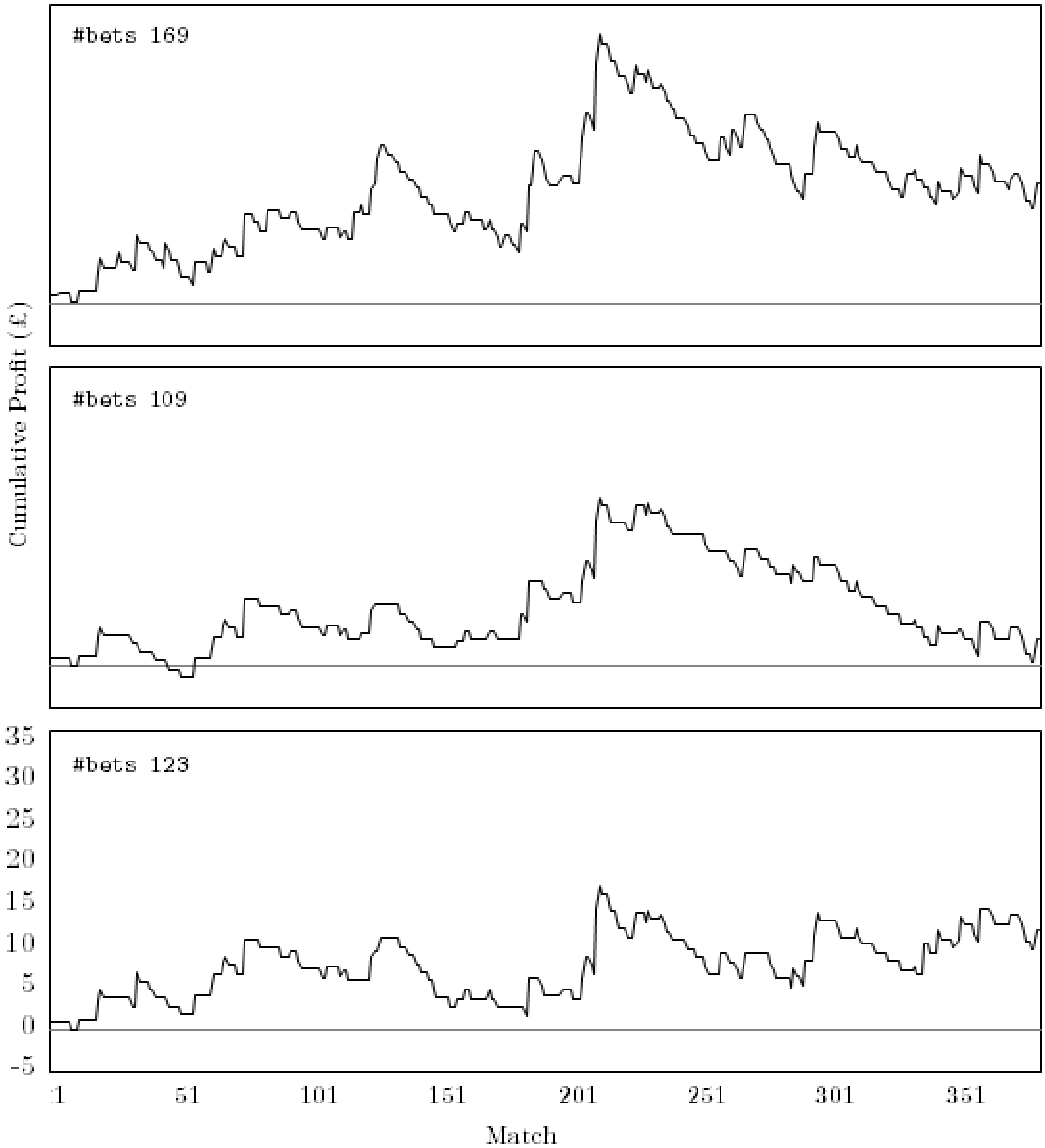


Figure 6.10. Cumulative profit/loss observed given f_S when simulating the standard betting strategy at discrepancy levels of $\geq 5\%$ against a) f_{maxB}, b) f_{meanB} and c) f_{WH}.

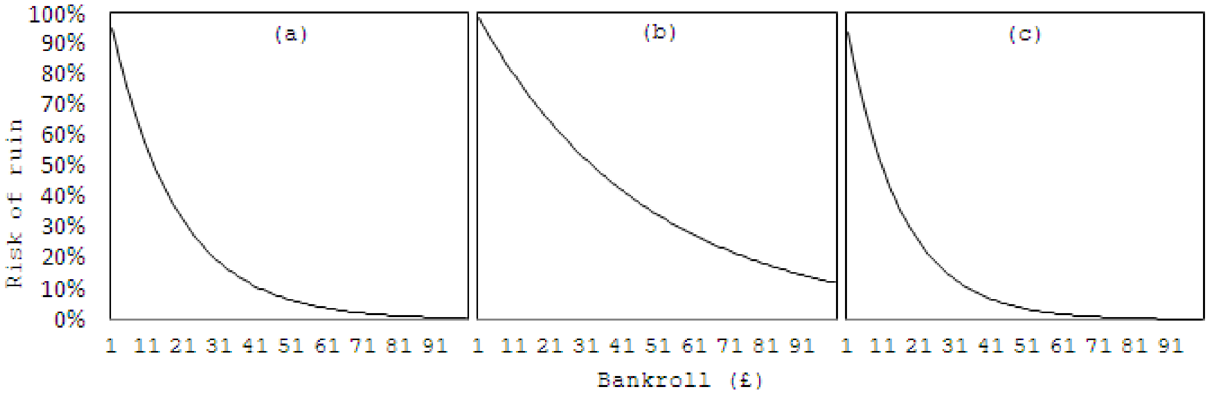


Figure 6.11. Risk of Ruin given the specified betting simulation against a) f_{maxB}, b) f_{meanB} and c) f_{WH}.

Table 6.2. Betting simulation stats given f_S against) f_{maxB}, b) f_{meanB} and c) f_{WH} at discrepancy levels of ≥ 5%.

	f_{maxB}	f_{meanB}	f_{WH}
Total bets	169	109	123
Bets won	57 (33.72%)	38 (34.86%)	44 (35.77%)
Total returns	£183.19	£112.13	£134.66
Min. P/L balance observed	£0.28	-£0.04	-£0.09
Max. P/L balance observed	£30.67	£19.86	£16.86
Final P/L balance	£14.19	£3.13	£11.66
Profit/Loss (%)	8.40%	2.87%	9.48%
Max. bookmakers considered per instance	40	40	1
Min. bookmakers considered per instance	28	28	1
Mean bookmakers considered per instance	35.73	35.73	1
Max. odds won	9	7.73	8.5
Min. odds won	1.19	1.40	1.40
Mean odds won	3.21	2.95	3.06
Mean profit margin (for all 380 instances)	0.63%	6.09%	6.50%
Arbitrage instances (for all 380 instances)	62	0	0

Table B.6.1 provides further statistics when performing this betting simulation given f_S against f_{maxB}, f_{meanB}, and f_{WH} using discrepancy levels that are different from the standard 5%. In general, pi-football appears to perform much worse at the lowest discrepancy levels (from 1% to 3%) and much better at higher discrepancy levels (from 4% to 11%). Considering a minimum of 30 simulated bets, the maximum profits are observed at discrepancy levels of 11% (35.63%), 9% (8.86%) and 8% (10.07%) against f_{maxB}, f_{meanB}, and f_{WH} respectively. At discrepancy levels above 11% there were too few betting instances to be able to derive meaningful conclusions.

6.4.3. Impact analysis of the subjective components

Table 6.3 describes profitability and accuracy performances based on the specified combinations of active components (*fatigue*, *psychology*, and *form*) relative to prior performances given f_O. Here, a component state is assumed to be *true* for a given match instance when there is ≥ 5% absolute discrepancy between competing teams in subjective proximity for that component. For example (scenario 1) there were 40 matches in which the subjective proximity was ≥ 5% for all three components.

Table 6.3. Analysis of impact of different subjective components of the model on profitability and accuracy

Scenario	Component/s True	Relevant Occurrences (given comp. state)	Bets Simulated (given the betting strategy)	Revised profitability relative to f_O (cumulative profit increase/decrease)	Revised accuracy relative to f_O (cumulative profit increase/decrease)	Measures agree in the direction of the revision
1 (All components True)	All	40	22	+£16.56	-0.4823	TRUE
2 (Exactly two components True)	Fatigue, Psychology	31	13*	+£7.70*	-0.3394	TRUE*
	Fatigue, Form	9*	2*	-£2.60*	-0.2004*	FALSE*
	Psychology, Form	57	27	+£6.28	-0.4008	TRUE
3 (Exactly one component True)	Fatigue	32	20	+£15.25	+0.0277	FALSE
	Psychology	79	31	+£5.67	-0.1694	TRUE
	Form	53	29	-£11.60	+0.0569	TRUE

* sample size too small to contribute to conclusions.

Considering both profitability and accuracy measures, it appears that all three components have contributed significantly in increasing the forecasting capability of this model, but it is dangerous to formulate strong conclusions about individual component-based performances due to the low numbers of relevant occurrences under the various scenarios. There is weak evidence that, in terms of profitability based on the betting simulation specified in Section 6.4.2, the *Fatigue* component appears to provide the highest overall improvement[14] when active, followed by the *Psychology* component that demonstrates improvements under all scenarios for which is active. The *Form* component appears to provide declines in profitability under scenario $3c$. In contrast, the accuracy measure suggests that the *Psychology* component provided the highest reduction in error under all the scenarios for which is involved, whereas components *Fatigue* and *Form* appear to provide very similar error fluctuations for all respective sub-scenarios.

6.5 Concluding remarks and future work

We have presented a novel Bayesian network model called pi-football $(v1.32)$ that was used to generate the EPL match forecasts during season 2010/11. The model considers both objective and subjective information for prediction, in which time-dependent data is weighted using degrees of uncertainty. In particular, objective forecasts are generated first and revised afterwards according to subjective indicators. Because of the 'anonymous' underlying approach which generates predictions by only considering the strength of the

[14] Evidence of slight decline under scenario $3b$ are based only on two simulated bets.

two competing teams given results data and total points, the entire model is easily applicable to any other football league.

For assessing the performance of our model we have considered both accuracy and profitability measurements since earlier studies have shown conflicting conclusions between the two and suggested that both measurements should be considered. In (Dixon & Coles, 1997) the authors claimed 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, and (Graham & Stott, 2008) suggested that if such a work was particularly successful, it would not have been published. Ours is the first study to demonstrate profitability against all of the (available) published odds. Previous studies have only considered a single bookmaker, since we are the first to prove that the published odds of a single bookmaker cannot be representative of the overall market (Chapter 5.3). In fact, pi-football was able to generate profit against maximum, mean, and common bookmakers' odds, even allowing for the bookmakers' in-built profit margin.

We showed that subjective information improved the forecast capability of our model significantly, and the evidence of this study agree with other recent relevant published studies whereby the knowledge of experts or preference of decision makers is employed in diverse forecast domains in an attempt to increase forecast precision and decision making (Joseph et. al., 2006; Min et al., 2008; Fu & Yang, 2012; Masegosa & Moral, 2012; Salmeron & Papageorgiou, 2012; Xiong et. al., 2012). Our study also emphasises the importance of Bayesian networks, in which subjective information can both be represented and displayed without any particular effort. Because of the nature of subjective information, we have been publishing our forecasts online at www.pi-football.com (pi-football, 2010) prior to the start of each match (earlier studies which incorporated subjective information have not done so). Appendix B.7 provides examples of both objective (f_O) and subjective (f_S) forecasts for match instances at the beginning of the EPL season 2010/11. At standard discrepancy levels of 5% the profitability of this model ranges from 2.87% to 9.48%, whereas at higher discrepancy levels (8% to 11%) the maximum profit observed ranges from 8.86% to 35.63%, depending on the various bookmakers' odds considered. No other published work appears to be particularly successful at beating all of the various bookmakers' odds over a large period of time, which highlights the success of pi-football.

Clearly the real potential benefits of a model such as this are critically dependent on both the structure of the model and the knowledge of the expert. A perfect BN model would still fail to beat the bookmakers at their own game if the subjective expert inputs are inaccurate. Because of the weekly pressure to get all of the model predictions calculated and published online, there was inevitable inconsistency in the care and accuracy taken to consider all the subjective inputs for each match; in most cases the subjective inputs were provided by a member of the research team who is certainly not an expert on the English premier League. If the model were to be used by more informed experts we feel it would provide posterior beliefs of both higher precision and confidence.

Chapter 7 extents this research by attempting to both simplify and improve the forecasting capability of this model, and this extended model is assessed against the subsequent EPL season of 2011/12.

CHAPTER 7

Profiting from an Inefficient Association Football Gambling Market: Prediction, Risk and Uncertainty Using Bayesian Networks

The novel material introduced in this chapter comes from our paper submitted for publication (Constantinou et al., 2012b), and presents a BN model for forecasting football match outcomes that is based on (Constantinou et al., 2012a), but demonstrates reduced complexity along with even higher forecasting capability.

7.1 Introduction

In Chapter 6 we presented a Bayesian network model that was used to generate forecasts about the EPL matches during season 2010/11, by considering both objective and subjective information for prediction. Forecasts were published online at www.pi-football.com 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 introduced profit margin.

In this chapter we present a Bayesian network model for forecasting football outcomes that is based on the approach in Chapter 6, but with reduced complexity and higher forecasting capability (which we explain in detail in Sections 7.2, 7.3 and 7.4). 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 match forecasts for the English Premier League (EPL) matches of season 2011/12, and forecasts were also published online at www.pi-football.com prior to the start of each match. Profitability, risk and uncertainty are evaluated by considering various unit-based betting procedures against published market odds. Overall, the model is able to generate even more profitable returns than the previously published model.

The chapter is organised as follows: Section 7.2 describes the model, Section 7.3 presents the various betting procedures along with a Bayesian network component for assessing the risks involved under each procedure, Section 7.4 discusses the results and Section 7.5 provides our concluding remarks.

7.2 The model

In this section we first provide a brief overview of the model summarising the main differences to the approach in Chapter 6. We then describe the technical components of the model in subsections.

As in Chapter 6 we have used the AgenaRisk Bayesian network tool to build the model. The model is constructed on the basis of three generic factors *(team strength, form, and motivation/fatigue)* and there are model components corresponding to each of the 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 allow us determine the effectiveness of each model component by assessing the probability distributions generated at each level of hierarchy. We reason with regards to the proposed component hierarchy as follows:

1. At level 1, match forecasts of type $\{p(H), p(D), p(A)\}$[1] are generated based on each team's strength (S), where an S prior is formulated according to a) observed and expected results (P) of relevant match instances of the current season, and b) team inconsistencies (I) given relevant final league points 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);

[1] Corresponding to home win, draw, and away win.

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 $\sim TNormal(\mu, \sigma, 0, 114)$[2] probability density function whereby posterior predictive distributions are formulated at each level of hierarchy which predict the strength of a team in total league points for the upcoming match. Distribution S is summarised in 14 predetermined ranks (S_R) as presented in Table 7.1, 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[3]; implying that the underlying 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 (MU) and Newcastle United (NU), and assuming that $S = 85(S_R = 1)$ and $S = 62(S_R = 6)$ respectively, the resulting forecasts will represent: "*a team with a probability density function $S(SR)$ whereby the maximum likelihood estimation is 85(1) plays against a team with a probability density function $S(S_R)$ whereby the maximum likelihood estimation is 62(6)*" instead of: "*Man United plays against Newcastle*". Accordingly, a team's S distributions vary

[2] 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 3 points for a win).

[3] The database consists of the home, draw and away results of all the EPL matches from season 1993/94 to 2010/11 inclusive (a total of 6624 occurrences). This information is available online at (Football-Data, 2012).

throughout the season, and it is possible for teams to share similar such distributions at certain periods throughout the season.

Table 7.1. How $S \to S_R$ is defined in 14 predetermined ranks (same as in Chapter 6).

S	>84	80-84	75-79	70-74	65-69	...(intervals of 5 points)	30-34	25-29	>25
S_R	1	2	3	4	5	...	12	13	14

Figure 7.1 illustrates a simplified model topology of the overall Bayesian network model and demonstrates how match forecasts transform on the basis of hierarchical posterior predictive distributions of S beliefs. Figure 7.2 presents the actual outcomes of the *Arsenal vs. Liverpool* EPL match as forecasted on August 20th 2011. The observed outcome was A (score was 0-2).

The primary differences with the BN model proposed in Chapter 6 are:

a) the model considers three generic factors that are inferred hierarchically in order to introduce differences between their significance and easy computation (instead of four averaged generic factors);
b) P, which formulated the prior predictive distribution of S, is measured using a straightforward $Beta - Binomial$ approach, rather than the non-symmetric Bayesian parameter learning approach;
c) model components which correspond to each of the generic factors have been simplified in an attempt to reduce model complexity;
d) supplementary information to relevant historical outcomes formulate posterior S distributions for each team, rather than directly updating match forecasts on the basis of subjective proximity about one team having advantage over the other for a specific model component.

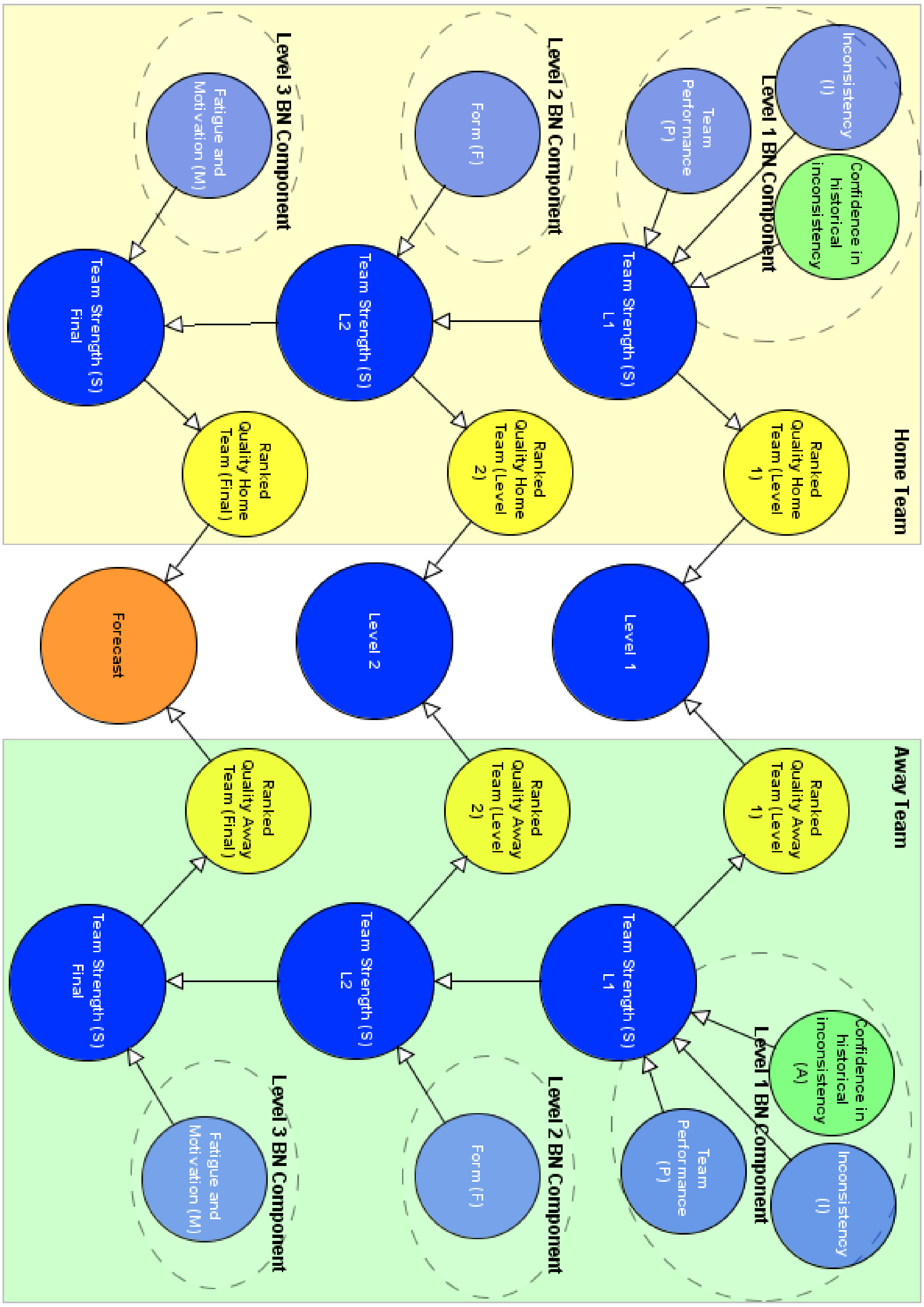


Figure 7.1. Model topology of the overall Bayesian network.

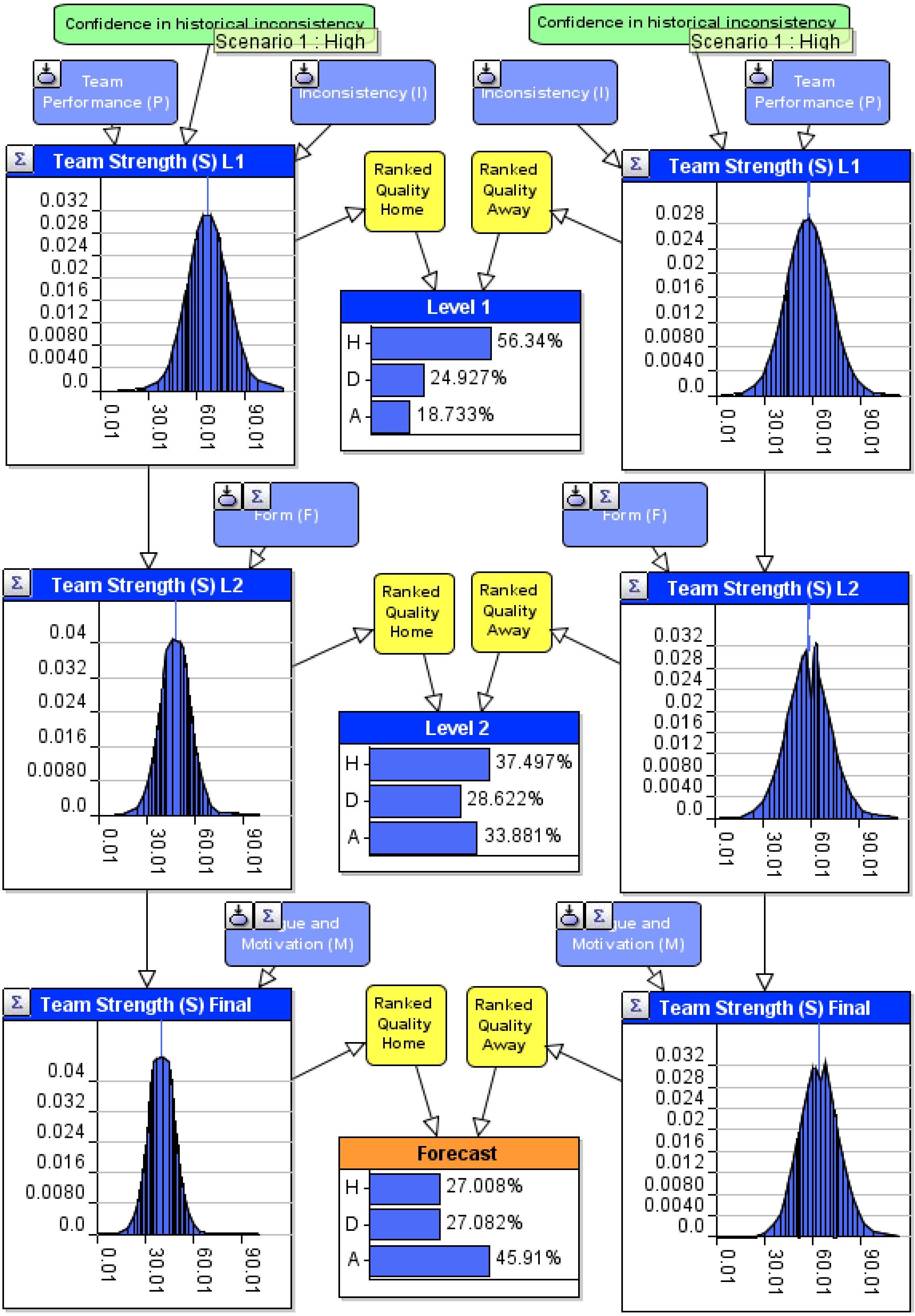


Figure 7.2. A simplified representation of the overall Bayesian network model. An example with the actual scenarios of the *Arsenal vs. Liverpool* EPL match, August 20th 2011. The observed outcome was $(0-2)$.

7.2.1. Level 1 Component: Team performance (P) and inconsistency (I)

At level 1, P is modelled using a $Beta - Binomial$ approach. In particular, $Beta$ distributions for each $\{p(W), p(D), p(L)\}$ beliefs are formulated by considering as hyperparameters the relevant previous season's observations, and formulate posterior beliefs which are inferred by $Binomial$ distributions with relevant observations from the current season. Consequently, for each $Beta$ distribution there exist $Binomial$ distributions that serve as the $alpha$ and $beta$ parameters[4].

The posterior $Beta$ distributions are then considered for formulating averaged expectations for the residual match instances of the current season, but expectations allow expert modifications based on subjective beliefs about the difficulty of residual opponents (this ensures against bias in cases where the current season results were only against poor quality or high quality teams). Observed and expected match points then formulate the prior distributions of S. This is the first part (out of two) of level 1. The Bayesian network component P is illustrated at Figure 7.3, where:

a) the variables Win, $Draw$ and $Lose$ are the posterior $Beta$ distributions. For example, in the case of $p(Win)$ the hyperparameters are $\sim Beta(w + 1, d + l + 2)$[5], where w is the number of wins during the previous season, d is the number of draws, l is the number of losses, and values 1 and 2 are introduced for minimal *Laplacian smoothing* so

[4] Effectively a multinomial distribution with beta distributions priors on each $\{p(W), p(D), p(L)\}$. The inputs will always ensure than p values $p(W) + p(D) + p(L) = 1$.

[5] Hyperparameters are provided as node-inputs and are not shown in Figure 7.3.

that we avoid overfitting by ensuring that posterior parameters $alpha$ and $beta$ are both positive for all teams;

b) the variables *number of wins, draws and loses* are $\sim Binomial(n, p)$. 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 (p is the $Beta$ distribution of $p(win)$ in this example);

c) 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 $W/D/L$[6];

d) 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 $\pm 10\%$ according to a 7-level subjective belief (the issue of choice of the subjective factor is discussed in Section 7.5), and it is defined as the case function of:

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

e) the variable *Current Points* simply represents the total number of points accumulated for the current season and hence, it is dependent on the relevant $Binomial$ observations.

[6] 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 lower levels.

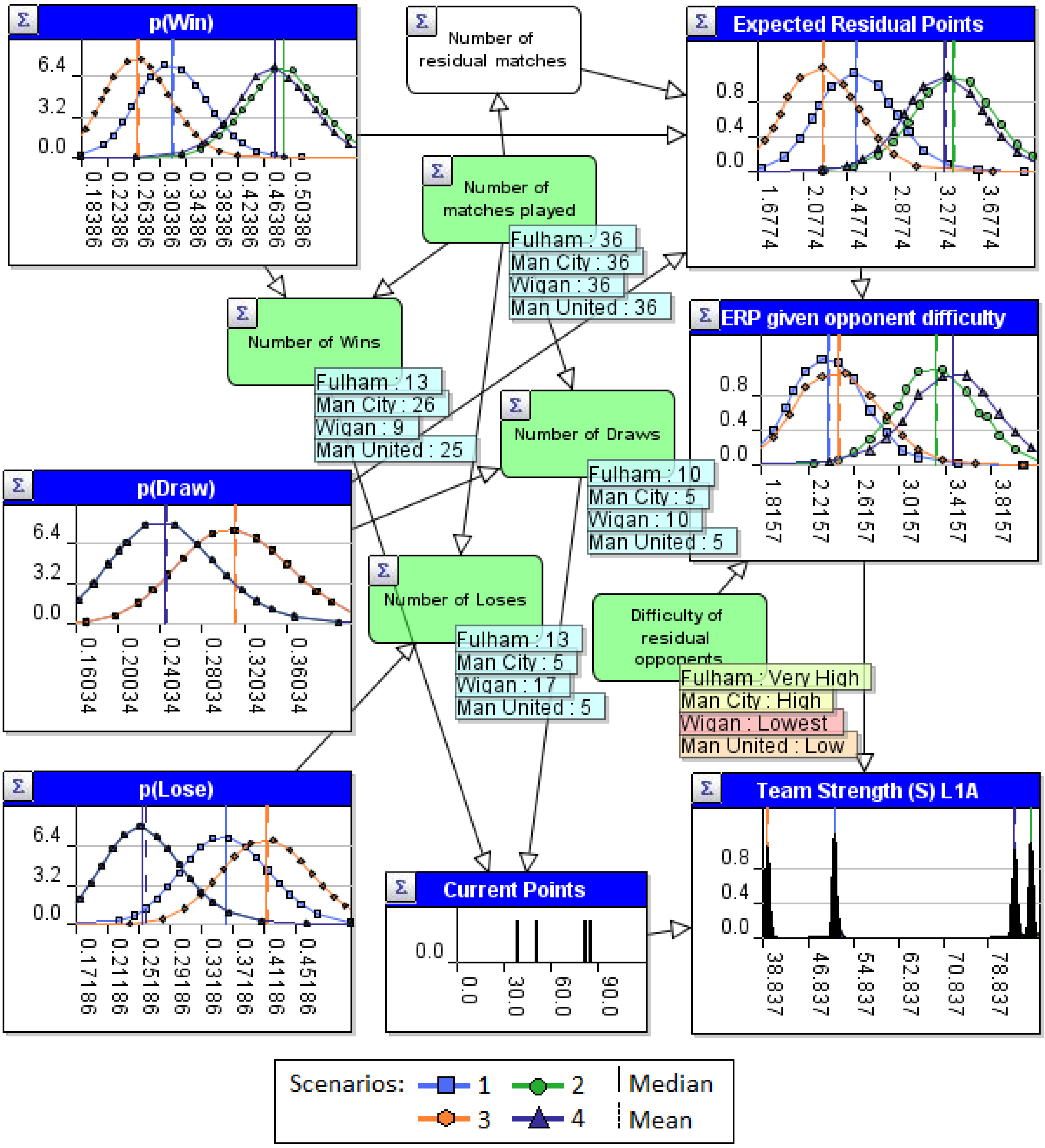


Figure 7.3. Level 1 Component (P): formulating the S prior. An example with four actual scenarios of Fulham, Man City, Wigan, and Man United, at gameweek[7] 37 during season 2011/12.

[7] A complete EPL season consists of 38 gameweeks.

The component inconsistency (I) approximates a team's inconsistency based on respective concluding league points over the five most recent seasons, and the resulting variance is added to the prior predictive distribution of S and together formulate S_{L1}. But the expert may avoid introducing additional variance if he or she feels that the team is not currently inconsistent. Figure 7.4 presents a naive parameter learning procedure for approximating a team's inconsistency, where:

a) the variables *Season* $Y1$ to $Y5$ are $\sim TNormal(\mu, \sigma, 0, 114)$;
b) the variable *Inconsistency* is a $\sim Uniform(0, 150)$[8] and is the variance (V) of the $TNormal$ distributions from (a);
c) the variable *Overall Performance* is a $\sim Uniform(0, 114)$ and is the mean of the $TNormal$ distributions from (a).

Figures 7.1 and 7.2 (from the previous section) present how the parts P and I of the level 1 component are connected, where the variable *Confidence in Historical Inconsistency* (C) is an ordinal scale distribution with subjective indications that allow the expert to reduce variance (I) additional to (S_{L1}) for up to 66.66% as the case function demonstrates below:

$$S_{L1} = \begin{cases} \sim TNormal\left(\pi_e, \frac{V}{3}, 0, 114\right), & \pi_e, c = Low \\ \sim TNormal\left(\pi_e, \frac{V}{2}, 0, 114\right), & \pi_e, c = Medium \\ \sim TNormal(\pi_e, V, 0, 114), & \pi_e, c = High \end{cases}$$

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

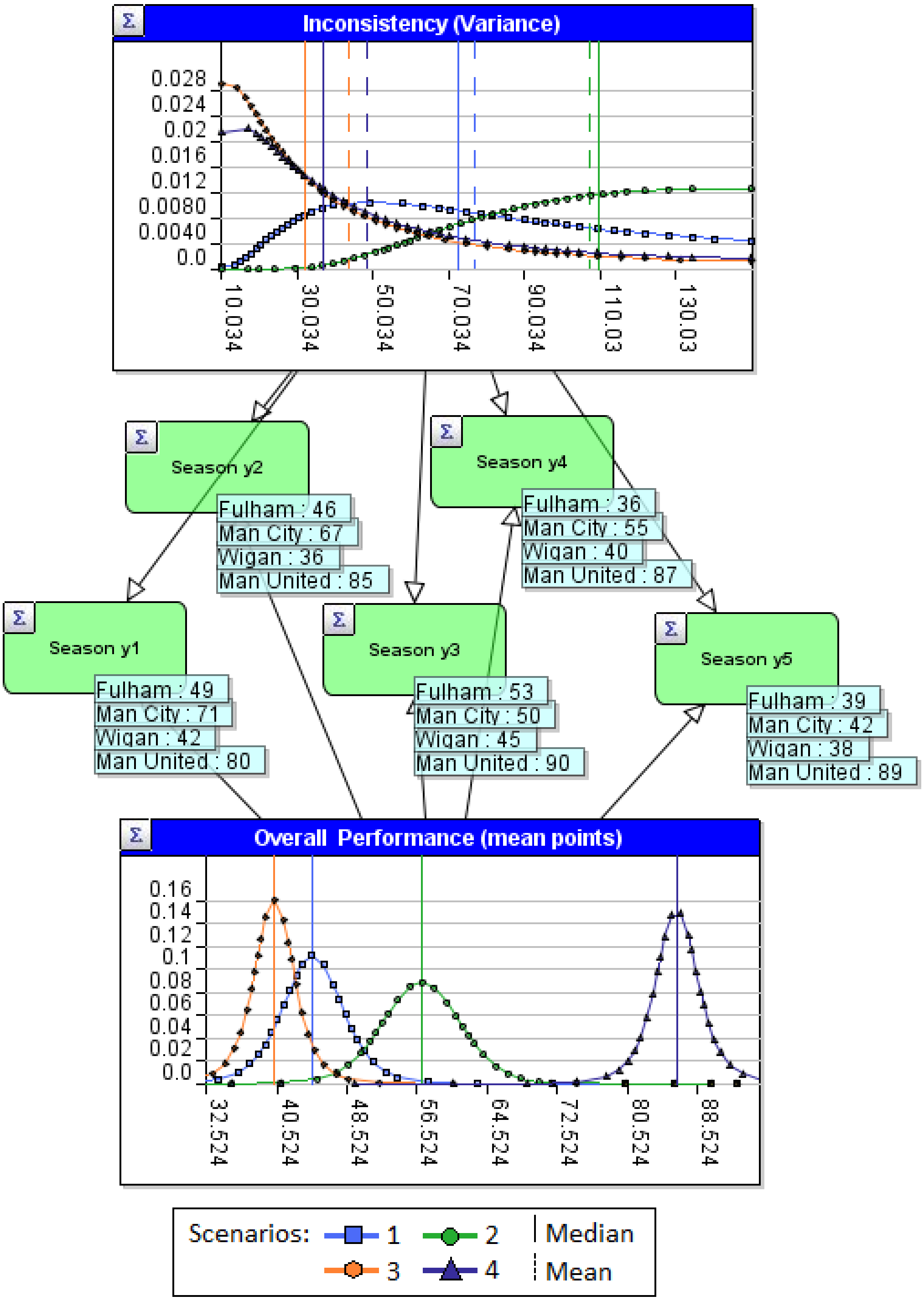


Figure 7.4. Level 1 Component (I): measuring a team's historical inconsistency (I) based on the concluding league points of the five most recent seasons. An example with four actual scenarios of Fulham, Man City, Wigan and Man United during season 2011/12.

7.2.2. Level 2 Component: Team form (F)

At level 2 a posterior predictive distribution S_{L2} is formulated given S_{L1} and a posterior team-form (Φ) value, as presented at Figure 7.5, 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. The expectations are determined according to model's math forecasts over the five most recent gameweeks. 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 by Figure 7.5 and the case functions below. The variable *Form given* L_A (Φ_{LA}) is the case function:

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

and the variable Φ_{LA} *Form given* L_R (Φ_{LR}) is the case function:

$$\Phi L_R = \begin{cases} TNormal(\Phi L_A, 0.01, 0, 1), & \Phi L_A, L_R = None \\ \sim TNormal\Big(\big(\Phi L_A + \big((1 - \Phi L_A) \times 0.1\big)\big), 0.01, 0, 1\Big), & \Phi L_A, L_R = Low \\ \sim TNormal\Big(\big(\Phi L_A + \big((1 - \Phi L_A) \times 0.2\big)\big), 0.01, 0, 1\Big), & \Phi L_A, L_R = Medium \\ \sim TNormal\Big(\big(\Phi L_A + \big((1 - \Phi L_A) \times 0.3\big)\big), 0.01, 0, 1\Big), & \Phi L_A, L_R = High \end{cases}$$

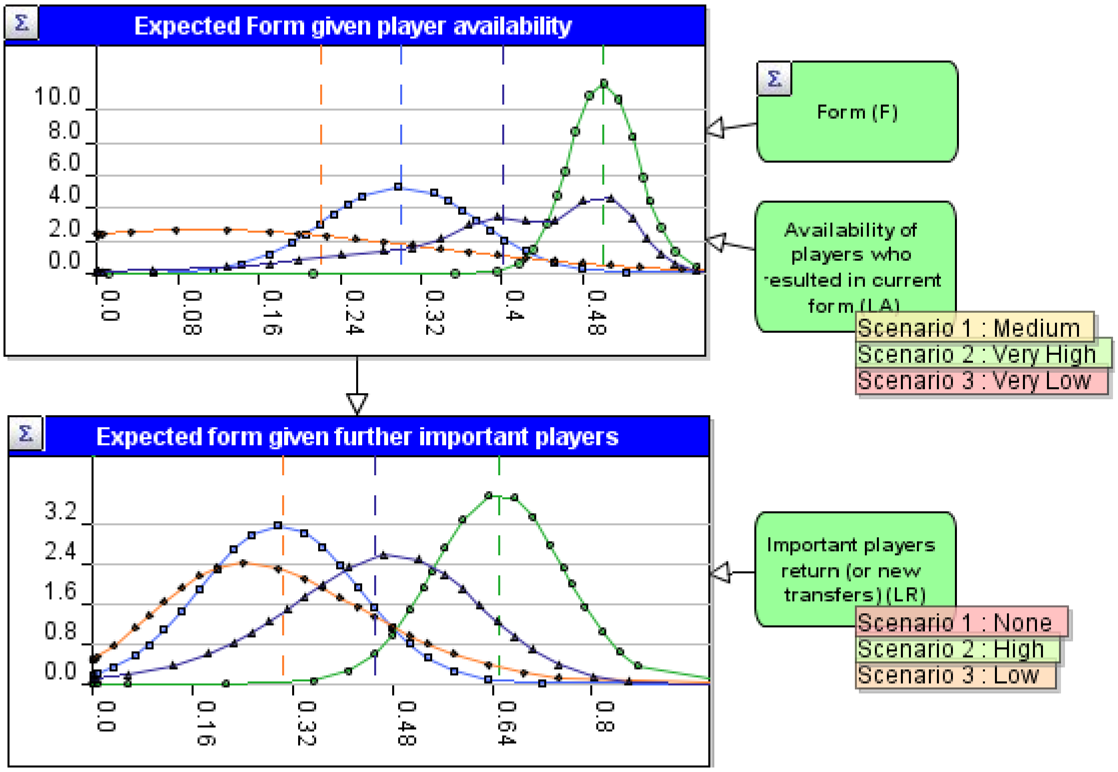


Figure 7.5. Level 2 Component (F): measuring team form. An example with four scenarios (scenario 4 represents uncertain inputs whereby values follow predetermined subjective probabilities).

7.2.3. Level 3 Component: Fatigue and motivation (M)

At level 3 a posterior predictive distribution of S_{L3} is formulated given S_{L2} and team fatigue and motivation as presented at Figure 7.6. A *Prior Fatigue* (G_p) is first measured according to the *EU match Involvement* (E) (which means involvement in an intermediate European tournament match) and the *Toughness of previous match* (T), where E and T follow ordinal scale distributions with subjective indications as illustrated by Figure 7.6 and the case function below. G_p is the case function:

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

The *Expected Fatigue* (G_e) is a posterior G_p value which diminishes on the basis of the *Days Gap since previous match* (δ), and increases with the *National Team Involvement* (λ), where δ and λ are ordinal scale distributions with subjective indications as illustrated by Figure 7.6 and the case function below. G_e is the case function:

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

The concluding variable G is measured given the *Motivation* (κ) and the *Head-to-Head Bias* (ω), where κ and ω follow ordinal scale distributions that go from 0 to 1 with subjective indications as illustrated by Figure 7.6 and the case function below. G is the case function:

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

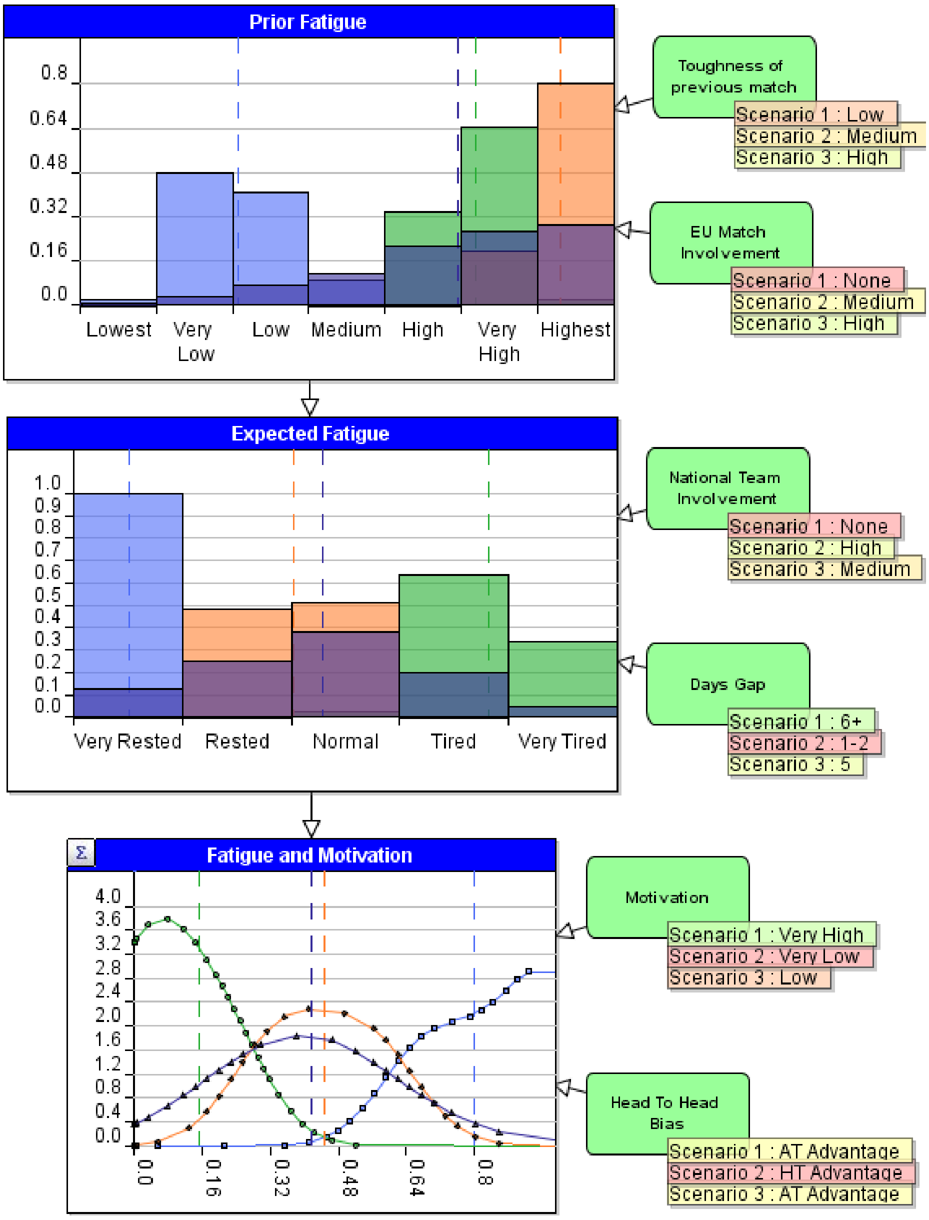


Figure 7.6. Component 3 (M): measuring fatigue and motivation. An example with four scenarios (scenario 4 represents uncertain inputs whereby values follow predetermined subjective probabilities).

7.3 Forecast performance based on profitability and risk

In this section we describe how the forecast capability of the model was assessed on the basis of profitability and relevant risks involved, and according to a set of predetermined betting procedures. However, for a profitability assessment to be possible the datasets of respective market odds are required. We have, therefore, considered the market odds with the highest payoff as recorded by (Football-Data, 2012) for the relevant matches of the EPL season 2011/12. The number of bookmaking firms considered for recording maximums ranged from 26 to 49 per match instance[9].

Naturally, the performance of football forecast models 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 (Dixon & Coles, 1997; Rue & Salvesen, 2000; Hirotsu & Wright, 2003; Goddard, 2005; Karlis & Ntzoufras, 2003; Goddard, 2005; Forrest et al., 2005; Joseph et al., 2006; Graham & Stott, 2008; Hvattum & Arntzen, 2010). 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 (see Chapter 4). Furthermore, in

[9] 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 (with highest payoff) a bettor can find. However, unlike traditional bookmakers Betfair will deduct a fixed % from your winnings which ranges from 2% to 6% depending on membership status (Betfair, 2000).

financial domains researchers have already demonstrated a weak relationship between various accuracy and profit measures (Leitch & Tanner, 1991), whereas (Wing et al., 2007) suggested that it might be best to combine accuracy and profit measures for a more informative picture.

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[10], where present, can be overcome. 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 previous relevant published papers (Pope & Peel, 1989; Dixon & Coles, 1997; Rue & Salvesen, 2000; Dixon & Pope, 2004; Goddard & Asimakopoulos, 2004; Forrest et al., 2005; Graham & Stott, 2008; Hvattum & Arntzen, 2010), whereby a bet is placed on outcomes for which the ratio of model to bookmakers' probabilities exceeds a predetermined critical level (also similar to that used in Chapter 6).

7.3.1. Defining profitability

In this chapter we measure the success of profitability on the basis of the quantity of profit (or net profit stated as unit-based returns), rather than on

[10] The bookmakers' profit margin, sometimes also called as 'over-round', refers to the margin by which the sum of the probability market odds of the total outcomes exceeds 1 by publishing odds with lower payoff than actual measured odds (higher in probability than actual measured probabilities) and thus, making the odds unfair for bettors (Chapter 5.3).

the basis of percentage returns relative to the respective stakes. The following example illustrates the rationale behind of this preference: 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 7.2 presents a hypothetical betting performance between the two models over those match instances.

Table 7.2. Hypothetical betting performance on the basis of profitability between the two football forecast models.

	α				β			
Match Instance	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%*	£500	£800	+£300	60%*

*Profit rate based on total stakes.

After considering the five match instances we observe the following results[11]:

- 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.

[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 less skilled at picking winners than model α. That is, such an evaluation fails to consider the possibility that model α might have failed to discover potential advantages (that other models have) for all of the match instances, and hence that model β managed to simulate riskier bets that reduced the percentage rates, 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, 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 at every match instance rather than choosing the 'best' of his advantages that occasionally arise.

Accordingly, we measure profitability on unit-based returns (net profit) over a simulated period over n match instances (which in our case $n = 380$ and represents the outcome over the whole EPL season of 2011/12). The betting procedures are defined in the following section.

7.3.2. 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, of which the model predicts with 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* for each 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 and it is equal to the *absolute* discrepancy percentage between outcomes multiplied by U (i.e. 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 and it is equal to the *relative* discrepancy percentage between outcomes multiplied by U (i.e. if an *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. ($BP_{5.1}$, $BP_{5.2}$, $BP_{5.3}$, $BP_{5.4}$,): Repeat 1, 2, 3 and 4 but substitute the betting procedure with arbitrage bets whereby the total amount of the three bets for the found instance is equal to the bankroll 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 potential events). Accordingly, we have constructed a simple Bayesian network component that measures the risk of ending with less than, or equal to, a specified number of units over a specified number of match instances. Figure 7.7 presents this network component and illustrates, as an example, the risk of ending with $U \leq 0$ after 380 match instances are simulated given BP_1 at discrepancy levels of 0%; assuming relevant model performances as demonstrated afterwards in Section 7.4. In particular, :

a) the variable *Match Instances* represents the number of match instances over which the risk is measured;
b) the variables *profitable* and *unprofitable* are $Beta$ distributions with $alpha$ and $beta$ parameters representing the probability to profit or not 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 and p probability of success equal to respective $Beta$ distributions 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 BP:1* 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.

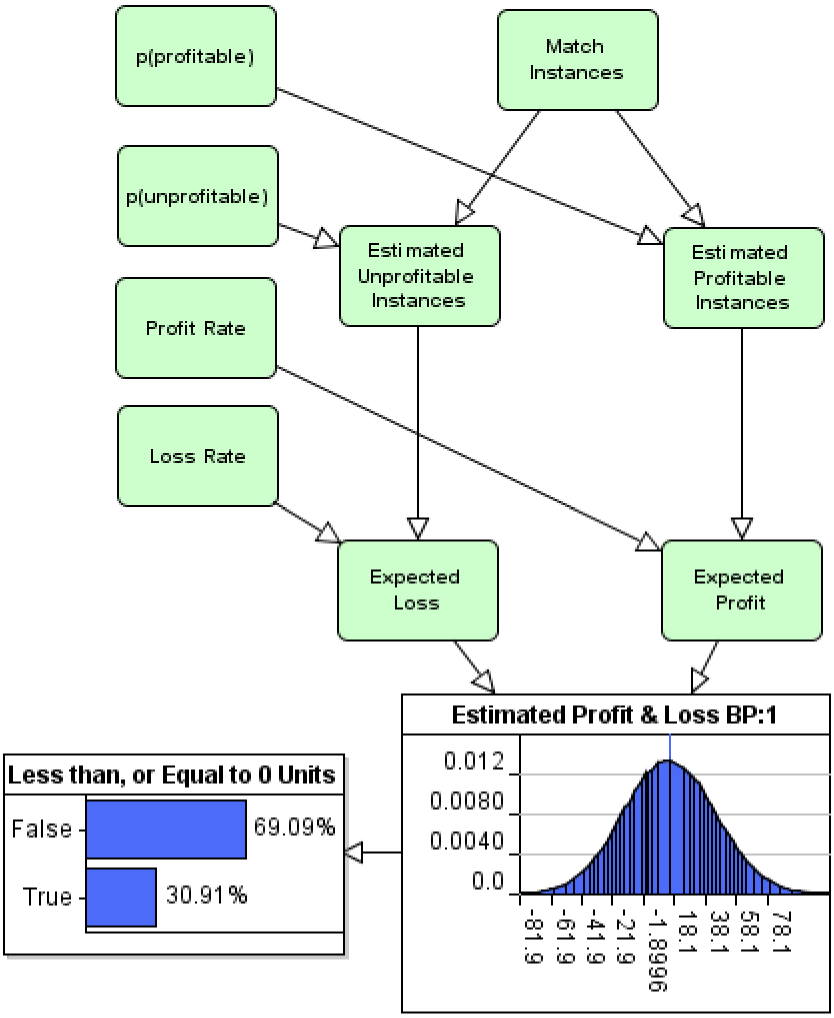


Figure 7.7. Bayesian network component for assessing the risks on final unit-based returns for each betting procedure.

7.4 Results and discussion

In this section we demonstrate and discuss the resulting performance of the model. In Section 7.4.1 we demonstrate the profitability of the model along with the relevant risks for each betting procedure; in Section 7.4.2 we evaluate the effectiveness of the model components based on transitions of profitability at each level of model hierarchy; in Section 7.4.3 we provide evidence of market inefficiency based on specific football teams; finally, in Section 7.4.4 we compare the performance of the model against the model in Chapter 6.

7.4.1. Model performance

Table 7.3 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. Figure 7.8 illustrates a graphical 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 the discrepancy levels of 6% and 1% respectively. In contrast, BP_2 generated maximum returns that are substantially higher relative to BP_1; returns of $U47.71$ and $U47.13$ at the discrepancy levels of 0% and 1% respectively. Figures C.1.1 and C.1.2 compare the cumulative returns over the season between the two betting procedures, whereby the results acknowledge that BP_2 is consistently generating 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 below.

As far as BP_1 and BP_2 are concerned, it is important to understand that we are much more confident for 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. But at higher discrepancy levels we require an even higher number of betting trials if we are to formulate conclusions that are as safe as those at the lower levels. To verify this, let us 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) 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, Figure 7.8 shows that an additional successful such 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) A plays B and A is a strong favourite, but stronger than bookies think. Consequently, the bookies offer a probability of 70% that team A wins. The model, however rates the probability as 82% and so we bet on team A to win. 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 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[12].

[12] 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 football but also in many different markets (Ali M., 1977; Quandt, 1986; Thaler & Ziemba, 1988; Shin H., 1991, Shin R. E., 1992; Shin H., 1993; Woodland & Woodland, 1994; Vaughn Williams & Paton, 1997; Golec & Tamarkin, 1998; Jullien & Salanie, 2000; Constantinou & Fenton, 2012b). Various theories exist, such as risk-loving behaviour, on why people are willing to bet on such uncertain propositions (Sobel & Raines, 2003; Snowberg, 2010).

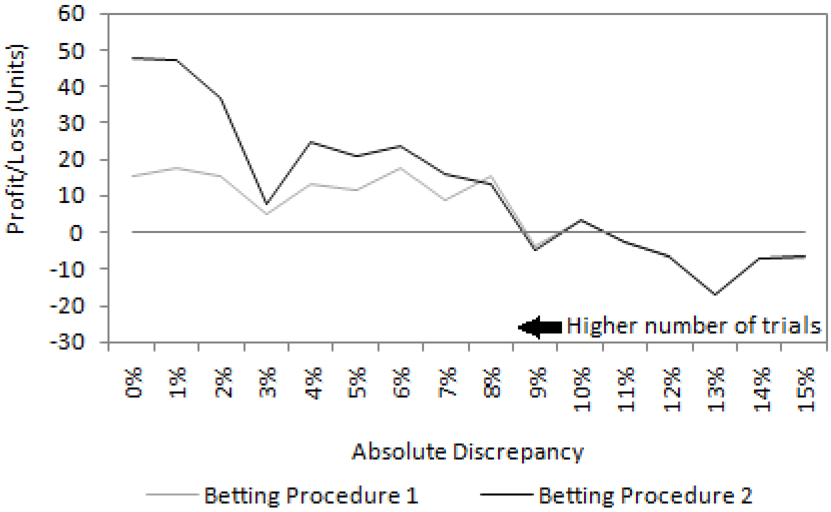


Figure 7.8. Concluding unit-based returns based on BP_1 and BP_2, and according to the specified level of discrepancy.

Table 7.3. Profitability and rates based on BP_1 and BP_2.

	Betting Procedure 1 (BP_1)				Betting Procedure 2 (BP_2)			
Discrep. levels (%)	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%

Figures 7.9 and 7.10 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 conlcuding balance of

BP_3 at match instance 380 is $U180.34$, whereas for BP_4 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 of Figures 7.9 and 7.10 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 $-U1{,}066.33$ (14.54% less relative to its maximum profit of $U1{,}247.86$). 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 7.4 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 Figure C.2.1 presents the respective predicted probability density risk distributions.

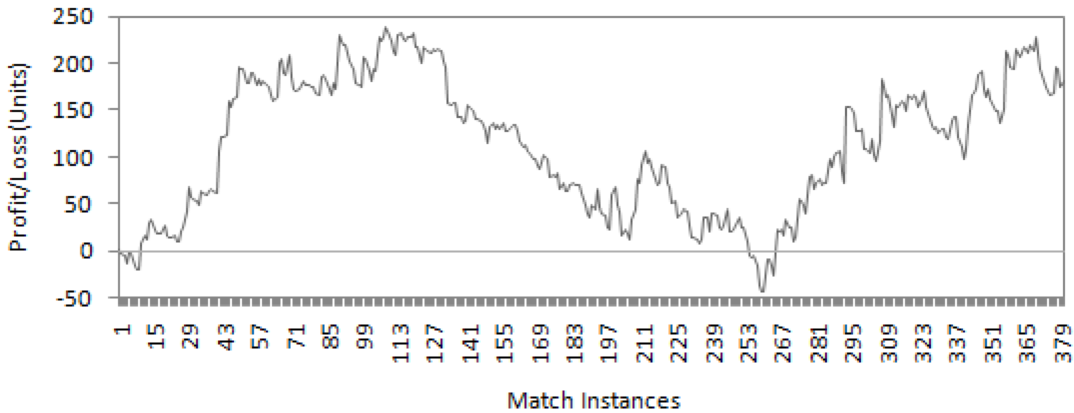


Figure 7.9. Cumulative unit-based returns based on BP_3.

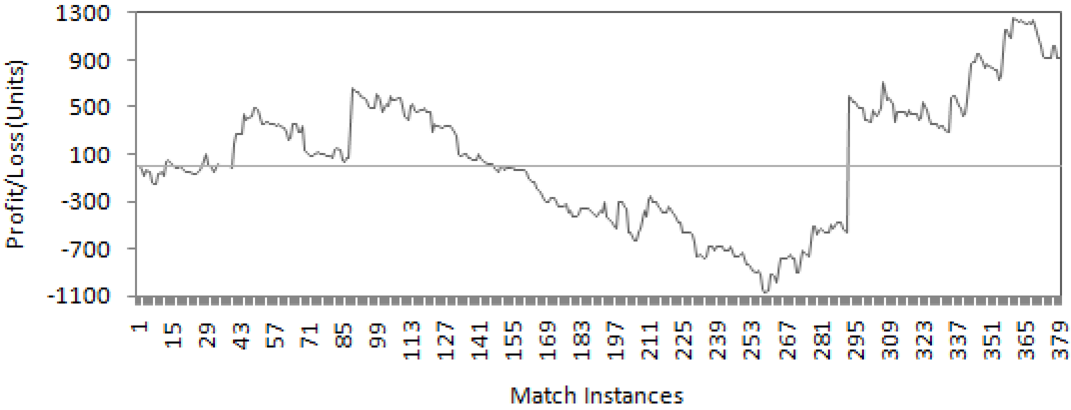


Figure 7.10. Cumulative unit-based returns based on BP_4.

7.4.2. Arbitrage opportunities and risk management

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. Figures C.3.1, C.3.2, C.3.3 and C.3.4 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 7.3.2). Table 7.4 and Figure C.2.1 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 risk free instances the profitability is heavily dependent on the initialised bankroll. In particular, bankrolls with sufficiently high initialised values (i.e. $\geq 1{,}000$ or $\geq 10{,}000$ in this case) eventually overshadow the predictive performance of the model since generated returns converge towards the arbitrage profit rate (since when an arbitrage opportunity is discovered the bet is equal to the value of the bankroll at that specific period).

Table 7.4. Risk probability values for the specified concluding returns[13] per betting procedure.

	Expected Profit/Loss (less than)								
BP	U1,000	U500	U100	U50	U0	-U50	-U100	-U500	-U1,000
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%

7.4.3. Effectiveness of model components

Figures 7.11, 7.12 and 7.13 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. Figures C.4.1, C.4.2, C.4.3, and C.4.4 demonstrate the profitability of the model over procedures BP_1, BP_2, BP_3 and BP_4 when:

[13] Results assume no discrepancy restrictions (set to 0%) for BP_1, BP_2, $BP_{5.1}$, $BP_{5.2}$, and an initialised bankroll of $10{,}000$ for the betting procedures of series 5.

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 finest), 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. This suggests that we have, most likely, overestimated the negative impact of *fatigue* for a team (e.g. the number of days gap since last competing match, the toughness of previous match, 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.

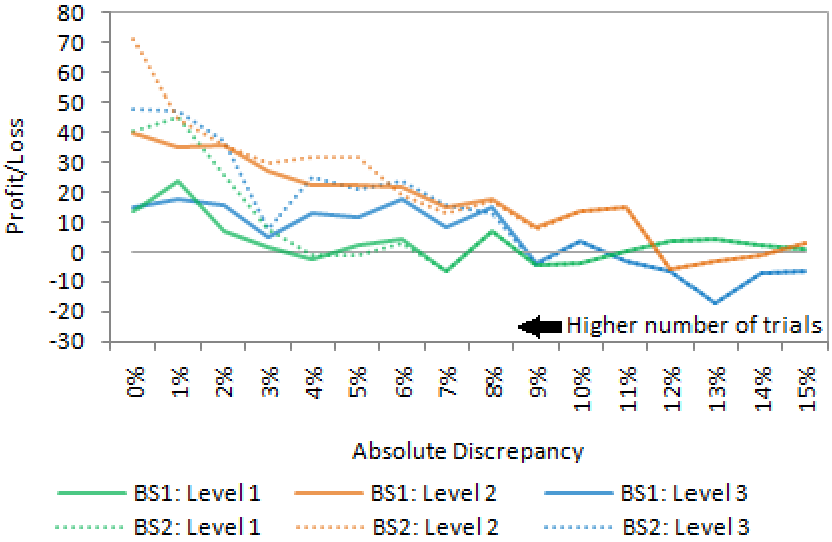


Figure 7.11. Cumulative unit-based returns based on BP_1 and BP_2, for component levels 1, 2 and 3.

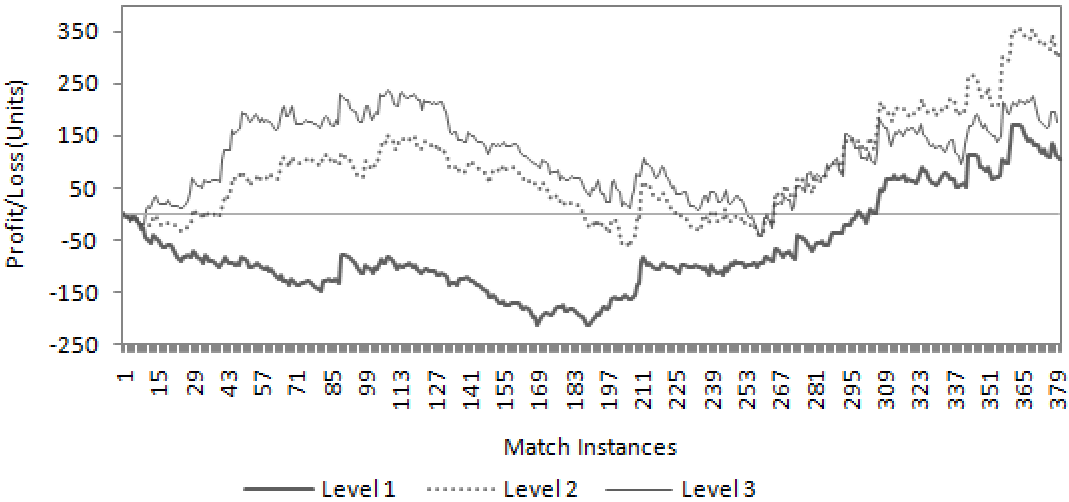


Figure 7.12. Cumulative unit-based returns based on BP_3, for component levels 1, 2 and 3.

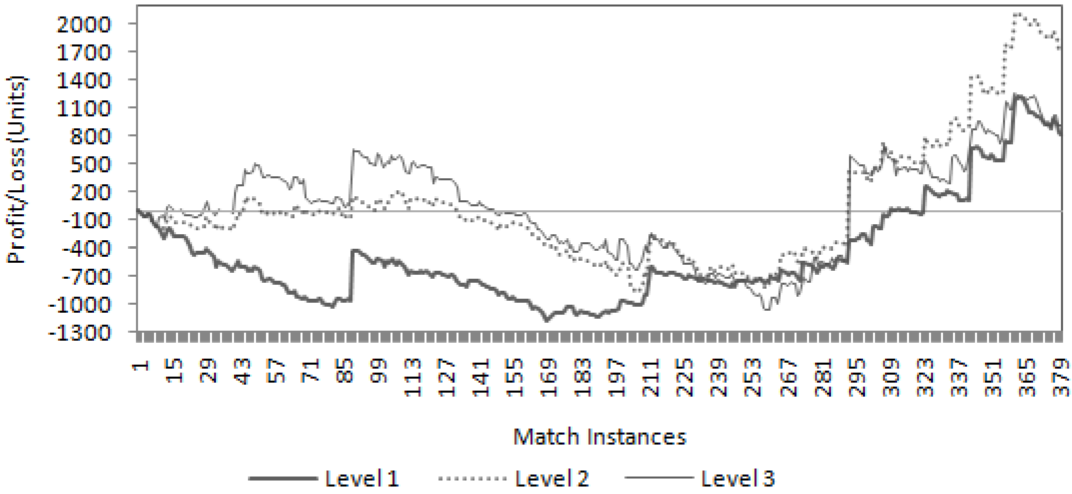


Figure 7.13. Cumulative unit-based returns based on BP_4, for component levels 1, 2 and 3.

7.4.4. 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 (Chapter 5). Table 7.5 demonstrates a team-based profitability assessment, 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 notable higher returns relative to the overall team, whereas for match instances involving Chelsea our model generated notable lower returns. Figure C.5.1 illustrates the team-

[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.

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 EPL 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 has been 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.

Table 7.5. Team-based returns relative to overall returns for the specified betting procedure.

		Betting Procedure:								
Rank	Team	1	2	3	4	5.1.	5.2.	5.3.	5.4.	Average
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%

7.4.5. Performance comparison against the previously published BN model

Figures C.6.1, C.6.2 and C.6.3 compare the unit-based cumulative returns over a period of 380 match instances (but for different seasons[15]) 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 model

[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 2011-12.

[16] Following the discussion in Section 7.4.1, we have ignored the scenarios whereby the discrepancy levels of BP_1 and BP_2 are set to $\geq 11\%$.

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 7.3.1, and Tables 7.3 and 7.6 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.

Table 7.6. Previous model's profitability based on BP_1 and BP_2 (for season 2010-2011).

	Betting Procedure 1 (BP_1)				Betting Procedure 2 (BP_2)			
Discrep. levels (%)	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%

7.5 Concluding remarks

We have presented a Bayesian network model for forecasting football match outcomes that not only simplifies a previously publish BN model, but also provides improved forecasting capability. The model considers both objective and subjective information for prediction, whereby subjective indications represent evidence that are important for prediction but which historical data fails to capture. The model was used to generate the match forecasts for the EPL season 2011/12, and forecasts were published online at www.pi-football.com prior to the start of each match.

For assessing the forecast capability of our model, we have introduced an array of betting procedures that are variants of a standard betting methodology that has been previously considered for assessing profitability by published relevant 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 superior to that generated by the old 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 replacing the specified betting procedure with arbitrage bets for match instances whereby odds between different firms guarantee risk free returns;

f) a team-based profitability assessment revealed further market inefficiencies (to the already extensive list) whereby published odds are consistently biased towards the *trademark* 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.

As in Chapter 6, also in this chapter the results not only emphasise the importance of Bayesian networks whereby subjective information can both be represented and displayed without any particular effort, but also how such belief networks can be used to enhance our understanding over uncertainty and our exposure to the relevant risks involved.

CHAPTER 8

pi-ratings: Determining the level of ability of football teams by dynamic ratings based on the relative discrepancies in scores between adversaries

The novel material introduced in this chapter comes from our paper submitted for publication (Constantinou & Fenton, 2012c), and proposes a novel and simple approach for dynamically rating football teams solely on the basis of the relative discrepancies in scores through relevant match instances. Even though the primary objective of this technique is rating and not prediction, the ratings can be incorporated in football forecast models such as those presented in Chapters 6 and 7, which solely focus on result-based outcomes for prediction, in an attempt to further enhance their forecasting capability.

8.1 Introduction

A rating system provides relative measures of superiority between adversaries. Determining the relative ability between adversaries is probably the most important element prior to football match prediction, and the current league positions are widely assumed to be an accurate indication of this. However, league positions suffer from numerous drawbacks which makes them unreliable for prediction. For instance, a football league suffers from high variation at the beginning of the season, and from low variation by the end of the season. Additionally, competing teams during a season might not share the equivalent number of matches played due to postponements and thus, the league table will be erroneous for many weeks. In fact, the league table is inherently biased until the final match of the season is played, because for the ranking to be 'fair' each team has to play against residual teams on home and away grounds. Even at the end of the season, the ranking represents the overall performance over the period of a whole season, and fails to demonstrate how the ability of a team varied during that period. Further, it ignores Cup matches and matches from other competitions (e.g. Champions League), and fails to compare teams in different divisions/leagues. In summary, a league table will never be a true indicator of a team's current ability at any specific time. A rating system should provide relative measures of superiority between adversaries and overcomes all of the above complications.

In Chapters 6 and 7 we demonstrated how some of the disadvantages concerning team performances based on league tables can be overcome by introducing further model parameters that reflect a team's form and hence, adjust the ability of a team according to the inconsistencies between expected and observed recent match performances. Furthermore, even though the models presented in Chapters 6 and 7 appeared to be particularly successful at beating bookmakers' odds, their forecasts did not incorporate score-based information about the relevant football teams.

In this chapter, we propose a novel and simple approach for dynamically rating association football teams solely on the basis of the relative discrepancies in scores through relevant match instances. This technique generates ratings that are meaningful in terms of diminished score difference against the average opponent and is applicable to any other sport where the score is considered as a good indicator for determining the relative performances between adversaries. In an attempt to examine how well the ratings captures a team’s performance, we have used them as the basis of a football betting strategy against published market odds and demonstrated profitability over a period of five English Premier League seasons (2007/08 to 2011/12), even allowing for the bookmakers' build-in profit margin. This is the first academic study to demonstrate profitability against market odds using such a simple technique. Even though the primary objective of this technique is rating and not prediction, the ratings can be incorporated in football forecast models such as those presented in Chapters 6 and 7, which solely focus on result-based outcomes for prediction, in an attempt to further enhance their forecasting capability.

This chapter is organised as follows: Section 8.2 presents the rating system, we discuss the results in Section 8.3 and we provide our concluding remarks and future work in Section 8.4.

8.2 The rating system

The rating system, which we call pi-rating, generates performance values that are meaningful in terms of diminished score difference (goals in this case) relative to the average opponent. A new team receives an initial rating of 0, and a rating of 0 represents the rating of the average team relative to the residual teams[1]. This implies that no inflations or deflations of overall ratings occur over time and thus, if one of the teams gains rating n then the adversary loses rating n.

When it comes to football, to generate ratings that accurately capture a team's current ability, we have to at least consider:

a) the well known phenomenon of *home advantage* (Clarke & Norman, 1995; Hirotsu & Wright, 2003; Poulter, 2009);

b) the fact that most recent results are more important than less recent when estimating current ability (see Chapters 6 and 7);

[1] If the rating is applied to a single league competition, the average team in that league will have a rating of 0. If the rating is applied to more than one league in which adversaries between the different leagues (or cup competitions) play against each other, the average team over all leagues will have a rating of 0.

c) the fact that a win is more important for a team than increasing goal difference.

In view of the above 'rules', we introduce the three following respective approaches:

a) different ratings for when a team is playing at home and away, and also a catch-up learning rate γ which determines to what extent the newly acquired information based on home performance influences a team's away rating and vice versa;

b) a learning rate λ which determines to what extent the newly acquired information of match goal-based results will override the old information in terms of rating;

c) for each individual match instance, a diminished reward ψ for each additional goal difference subsequent to 1 (for both predictions and observations).

8.2.1. Defining the pi-rating

When a team is playing at home then their new home rating is dependent on a) their current home rating, b) the opponent's current away rating, and c) the outcome of the match in terms of goal difference (and vice versa). In particular, the rating is developed in cumulative updates whereby diminished comparisons between expected and observed goal difference determines whether the rating will increase or decrease (i.e. a team's rating will increase if the score indicates a higher performance than that expected). Accordingly, the overall rating of a team is the average rating between home and away performances, and this is simply defined as:

$$R_\tau = \frac{R_{\tau H} + R_{\tau A}}{2}$$

where R_τ is the rating for team τ, $R_{\tau H}$ is the rating for team τ when playing at home, and $R_{\tau A}$ is the rating of team τ when playing away. Assuming a match between $\tau = \alpha$ and $\tau = \beta$, then the home and away ratings for teams α and β are respectively updated cumulatively as follows:

1. updating home team's home rating → $R'_{\alpha H} = R_{\alpha H} + \psi_H \times \lambda$
2. updating home team's away rating → $R'_{\alpha A} = R_{\alpha A} + (R'_{\alpha H} - R_{\alpha H}) \times \gamma$
3. updating away team's home rating → $R'_{\beta A} = R_{\beta A} + \psi_A \times \lambda$
4. updating away team's away rating → $R'_{\beta H} = R_{\beta H} + (R'_{\beta A} - R_{\beta A}) \times \gamma$

where $R_{\alpha H}$ and $R_{\alpha A}$ are the current home and away ratings of team α, $R_{\beta H}$ and $R_{\beta A}$ are the current home and away ratings of team β, $R'_{\alpha H}$, $R'_{\alpha A}$, $R'_{\beta H}$ and $R'_{\beta A}$ are the respective revised ratings, ψ is the diminished score difference between expected and observed performance (which we explain in detail in Section 8.2.2) and λ and γ are the learning rates (which we explain in detail in Section 8.2.3). Further, a step-by-step example of how the ratings are revised is presented in Section 8.2.4.

8.2.2. Diminishing rewards per additional score difference

Figure 8.1 demonstrates the diminishing rewards applied to each additional score difference greater than 1, for both observations (integer values) and predictions (real values). The difference between expected and observed goal difference is measured for the home and the away team respectively as follows:

$$\textit{Home Team} \rightarrow (g_H - g_A) - (e_{\Delta H} - e_{\Delta A})$$

$$\textit{Away Team} \rightarrow (g_A - g_H) - (e_{\Delta A} - e_{\Delta H})$$

where g_H and g_A are the number of goals scored by the home and away team respectively, and $e_{\Delta H}$ and $e_{\Delta A}$ are the expected goal difference values (non-diminished ratings) relative to the average adversary for the home and away team respectively. Consequently, the variable ψ will be the diminished value of the above formulations as presented in Table 8.1 and illustrated by Figure 8.1.

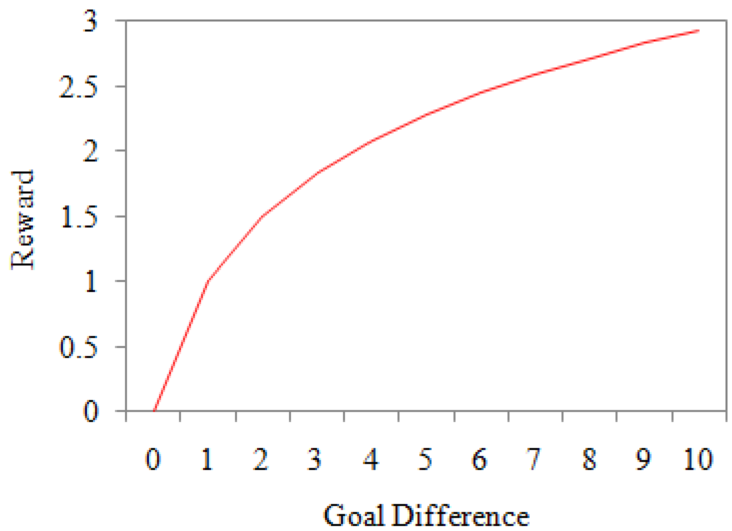


Figure 8.1. Cumulative diminishing rewards for subsequent goal difference values which are greater than 1.

Table 8.1. Distinct and cumulative diminishing returns for subsequent goal difference values which are greater than 1.

Goal Difference (GD)	Distinct Diminished Reward (GD^{-1})	Cumulative Diminished Reward (ψ)[2]
1	1	1
2	0.5	1.5
3	0.3333	1.8333
4	0.25	2.0833
5	0.2	2.2833
6	0.1666	2.45
7	0.1428	2.5928
8	0.125	2.7178
9	0.1111	2.8289
10	0.1	2.9289

8.2.3. Determining the learning rates

In football, new observations are always more important than the former, and no matter how home and away performances differ for a team, we can still gain some information about a team's next away performance based on its previous home performance (and vice versa). Thus, determining optimal learning rates for the variables λ and γ is paramount for generating ratings that accurately capture the current level of performance of a team.

The learning rates λ and γ can take values that go from 0 to 1. A higher learning rate λ determines to what extent the newly acquired information of match results will override the old information in terms of rating, and a higher learning rate γ determines the impact the home

[2] A linear diminished reward is introduced between two integer values (i.e. when the goal difference is set to $1 + \left(\frac{1}{2} \times 1\right) = 1.5$ then the cumulative diminished reward is $1 + \left(\frac{1}{2} \times 0.5\right) = 1.25$).

performances have on away ratings (and vice versa). For instance, when $\lambda = 0.1$ a team's rating will adjust with cumulative updates based on new match results with a weighing factor of 10%, and when $\gamma = 0.5$ a team's home performances will affect that team's away ratings with a weighting factor of 50% relative the revision of the home rating.

In determining the optimal learning rates we have assessed the ratings generated for different values of λ and γ by formulating score-based predictions about the last five English Premier League (EPL) seasons; 2007/08 to 2011/12. For training the learning rates[3] we have considered relevant historical data (Football-Data, 2012) beginning from season 1992/93 (and up to the previous season of that tested). Accordingly, if a combination of the learning rates λ and γ increase the forecast accuracy, then we assume that both λ and γ are a step closer to being optimal.

Figure 8.2 illustrates how parameters λ and γ affect the error in predicted score difference over the EPL seasons 1997/98 to 2006/07 inclusive, where the error is simply the difference between predicted and observed goal difference (e.g. if a model predicts $+1$ goal for the home team and the observation is $+1$ goal for the away team then the absolute score error is 2 goals). Our results show that combinations of λ and γ where $0.01 \leq \lambda \leq 0.02$ and $0.05 \leq \gamma \leq 0.7$ provide the best choices for optimum learning rates, and clearly demonstrate the significance of the λ parameter relative to γ.

[3] The first five EPL seasons (1992/93 to 1996/97) are solely considered for generating the initial ratings of the competing teams. This is important because training the model on ignorant team ratings (i.e. starting from 0) will negatively affect the training procedure. Thus, learning rates λ and γ are trained during the subsequent ten seasons; 1997/98 to 2006/07 inclusive.

Accordingly, we have chosen the learning rates of $\lambda = 0.02$ and $\gamma = 0.5$; values that are towards the maximum of those suggested as best choices by Figure 8.2 in order to allow for more rapid convergence of ratings in cases whereby teams spend hundreds of millions on new star players and completely change the profile of the team. This handles examples such as that of Manchester City's recent spending spree whereby a team that had failed to even challenge for the title in 44 years, improved so drastically in the last two years that they won it in 2012 (Scott M., 2012).

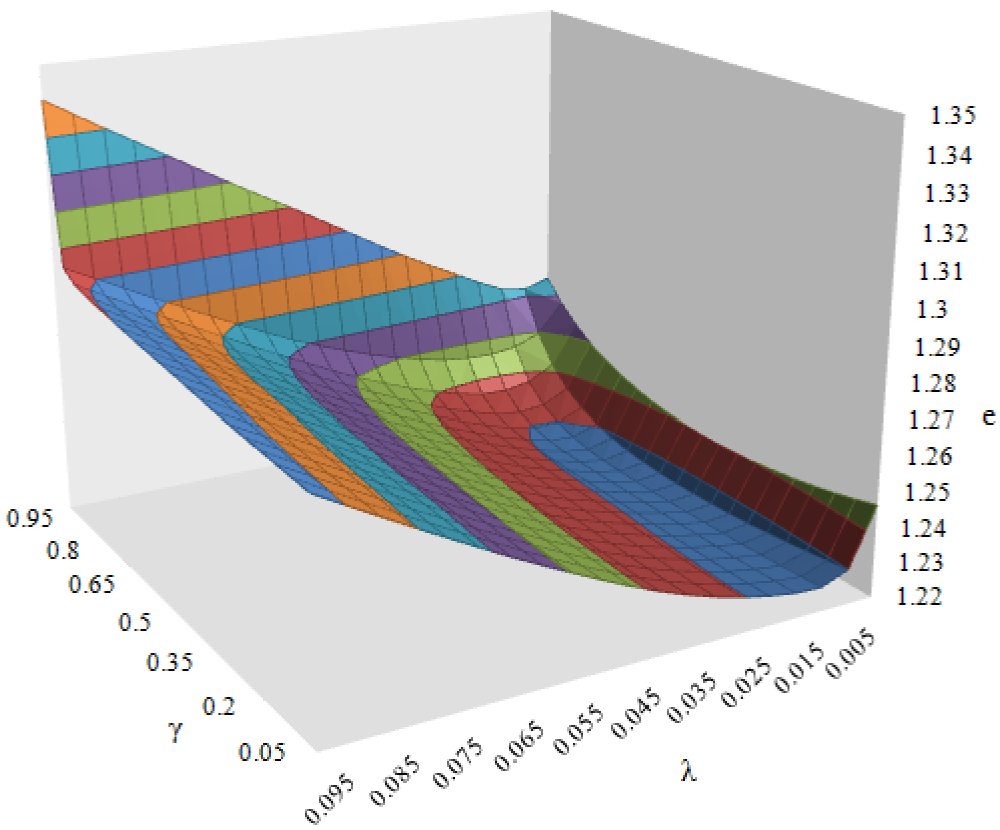


Figure 8.2. Estimating optimum λ and γ learning rates based on score-based error e for the EPL seasons 2007/08 to 2011/12.

8.2.4. Updating pi-ratings: An Example

Suppose that we have a match instance where team α (the home team) with ratings $\{R_{\alpha H} = 1.6, R_{\alpha A} = 0.4\}$ plays against team β (the away team) with ratings $\{R_{\beta H} = 0.3, R_{\beta A} = -1.2\}$. Converting the ratings to expected goal difference based on Figure 8.1 and Table 8.1 we retrieve the following information:

- team α is expected to win by 2.3 goals difference against the average opponent when playing at home;
- team α is expected to win by 0.4 goals difference against the average opponent when playing away;
- team β is expected to win by 0.3 goals difference against the average opponent when playing at home;
- team β is expected to lose by 1.4 goals difference against the average opponent when playing away.

Using the above information we can formulate predictions regarding the expected goal difference between the two teams at the specified ground. For this example, we have to consider team's α current home rating and team's β current away rating; the expected goal difference is $+3.7$ for team α. Suppose that we observe the score '$4-1$' ($+3$ for team α), and that the learning rates are set to $\lambda = 0.1$ and $\gamma = 0.3$. The old ratings are revised as follows:

Step 1: calculate the diminished rewards ψ based on the difference between expected and observed goal difference per team:

$$\text{team } \alpha \rightarrow (4-1) - \big(2.3 - (-1.4)\big) = 3 - 3.7 = -0.7^{4};$$

$$\text{team } \beta \rightarrow (1-4) - (-1.4 - 2.3) = -3 + 3.7 = +0.7.$$

Step 2: update team's α and team's β home and away ratings respectively based on the learning rate λ and variables ψ_H and ψ_A from step 1:

$$\text{team } \alpha \rightarrow \mathrm{R'_{\alpha H}} = 1.6 + (-0.7) \times 0.1 = 1.53 \text{ (down from } 1.6\text{)};$$

$$\text{team } \beta \rightarrow \mathrm{R'_{\beta A}} = -1.2 + (+0.7) \times 0.1 = -1.13 \text{ (up from } -1.2\text{)}.$$

Step 3: update team's α and team's β away and home ratings respectively based on the learning rate γ and revised ratings R'_{aH} and $R'_{\beta A}$ from step 2:

$$\mathrm{R'_{\alpha A}} = 0.4 + (1.53 - 1.6) \times 0.3 = 0.379 \text{ (down from } 0.4\text{)};$$

$$\mathrm{R'_{\beta H}} = 0.3 + \big(-1.13 - (-1.2)\big) \times 0.3 = 0.321 \text{ (up from } 0.3\text{)}.$$

Even though team α won team β '$4-1$', team's α ratings are decreased from $\{R_{aH} = 1.6, R_{aA} = 0.4\}$ to $\{R_{aH} = 1.53, R_{aA} = 0.379\}$, and team's β ratings are increased from $\{R_{\beta H} = 0.3, R_{\beta A} = -1.2\}$ to $\{R_{\beta H} = 0.321, R_{\beta A} = -1.13\}$. This happened because according to the ratings team α was expected to win team β by 3.7 goals.

[4] Since the difference is ≤ 1 the outcome is not diminished.

8.3 Betting performance and rating development

In an attempt to examine how well the rating captures a team's performance, we have used it as the basis of a football betting strategy against published market odds for the EPL seasons 2007/08 to 2011/12 inclusive. For assessing the profitability of the pi-rating system we have considered the learning rates $\lambda = 0.02$ and $\gamma = 0.5$, as suggested in Section 8.2.3, and formulated result-based predictions.

The predictions are based on how two adversaries with difference n in ratings performed throughout the training data; the ratings are divided into intervals of 0.10 (from $- \geq 1.1$ to $+ > 1.6$) and the closer the difference between ratings is to an interval the more important the historical predictive distribution of that interval becomes between the two[5]. The granularity of 28 intervals of team ratings has been chosen to ensure that for any rating combination (i.e. a team of rating x at home to a team of rating y) there are sufficient data points for a reasonably well informed prior for the result-based predictive distribution $\{p(H), p(D), p(A)\}$.

For betting simulation, we have followed a very simple strategy whereby for each match instance we place a £1 bet on the outcome with the

[5] The impact of the two intervals, for which the difference in rating between teams lies, is measured in absolute percentage difference from the rating value (i.e. if the rating value is 6 points away from interval x and 4 points away from interval y, then the impact of the predictive distribution of interval x is 60%, whereas it is 40% for that of interval y).

highest discrepancy of which the pi-rating system predicts with higher probability relative to published market odds[6].

Figures 8.3 and 8.4 demonstrate the distinct and overall cumulative profit/loss observed against published market odds during the five specified EPL seasons. Table 8.2 presents the summary statistics of the betting simulation. Overall, the technique is profitable which implies that the rating system properly captures the ability of a team at any time interval throughout the season.

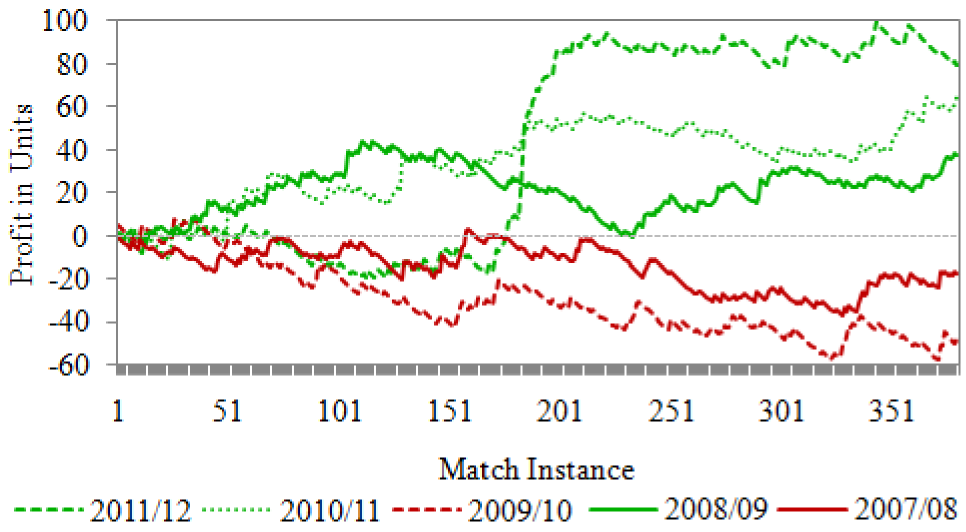


Figure 8.3. Distinct cumulative profit/loss observed against published market odds during the EPL seasons 2007/08 to 2011/12 inclusive.

[6] We have considered the *Betbrain* maximums (best available for the bettor) published odds as provided by (Football-Data, 2012) which are recorded on Friday afternoons.

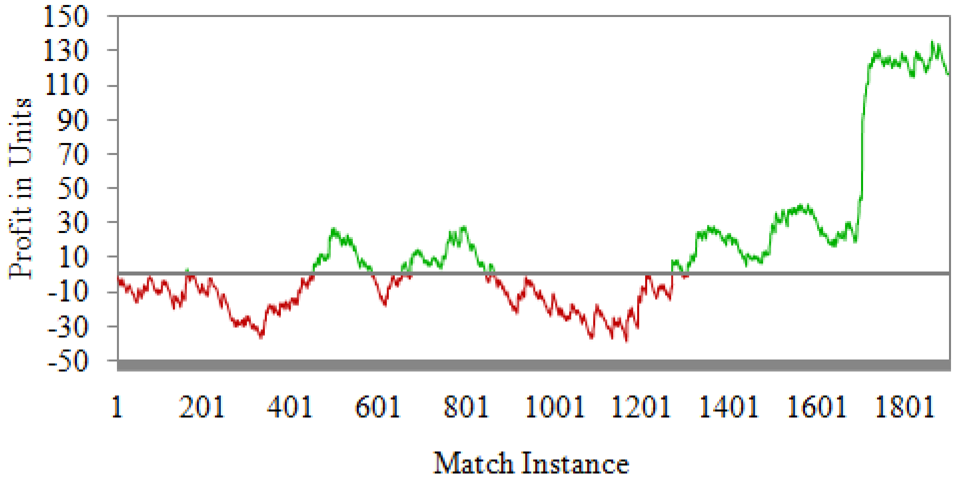


Figure 8.4. Overall cumulative profit/loss observed against published market odds during the EPL seasons 2007/08 to 2011/12 inclusive.

Table 8.2. Betting simulation: outcomes and statistics.

EPL season	Match instances	Number of bets	Bets won	Total stakes	Total returns	Profit/ Loss
2007/08	380	375	120 (32%)	£375	£357.56	-£17.44
2008/09	380	379	139 (36.67%)	£379	£416.23	+£37.23
2009/10	380	379	97 (25.59%)	£379	£330.52	-£48.48
2010/11	380	378	131 (34.65%)	£378	£442.85	+£64.85
2011/12	380	380	128 (33.68%)	£380	£459.82	+£79.82
TOTAL	1900	1891	615 (32.52%)	£1891	£2006.98	+£115.98

Figure D.1.1 illustrates how the pi-ratings develop for the six most popular EPL teams over the course of the last 20 seasons, whereas Figure 8.5 illustrates how the pi-ratings develop for those identical teams during the last five seasons (1900 match instances) if we consider no previous relevant historical information. In particular, at match instance 1 (first match of season 2007/08) all six teams start at rating 0. By considering the suggested learning rates of $\lambda = 0.02$ and $\gamma = 0.5$, the development of the rating shows that two seasons of relevant historical outcomes (76 match instances per team) might be enough for it to converge into acceptable estimates. However, a further season of historical match outcomes might be required for teams with the uppermost difference from the average team (i.e. Chelsea and Manchester United).

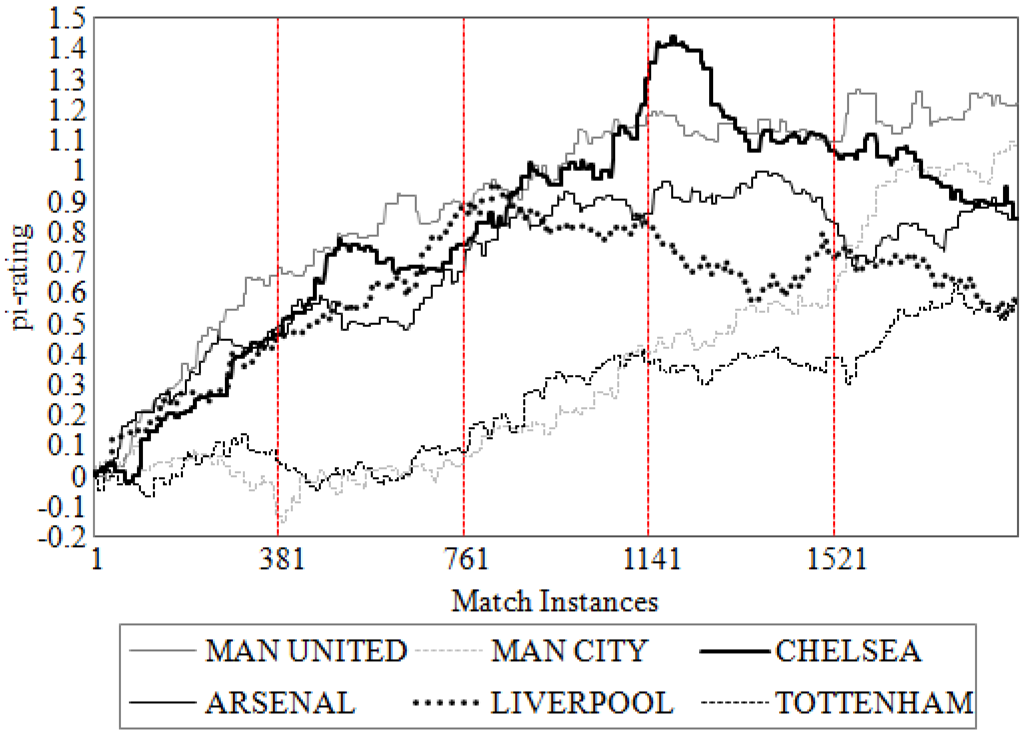


Figure 8.5. Development of the pi-ratings for seasons 2007/08 and 2011/12.

In contrast to earlier studies that assumed or concluded that the home advantage factor is invariant between football teams and hence considered a single generalised model parameter for that matter (Knorr-Held 1997, 2000; Koning, 2000; Baio & Blangiardo, 2010; Hvattum & Arntzen, 2010; Leitner, 2010), our results show that this is not the case. Figure 8.6 illustrates how the ratings develop on the basis of home and away performances for Manchester United, Blackburn, Wolves and Everton during the same five EPL seasons. In particular, Manchester United and Blackburn demonstrate a high variation between home and away performances, whereas Wolves and Everton appear to perform indifferently to home and away grounds. This outcome is consistent with (Clarke & Norman, 1995) who, in fact, reported that in many cases a team can develop a negative home advantage.

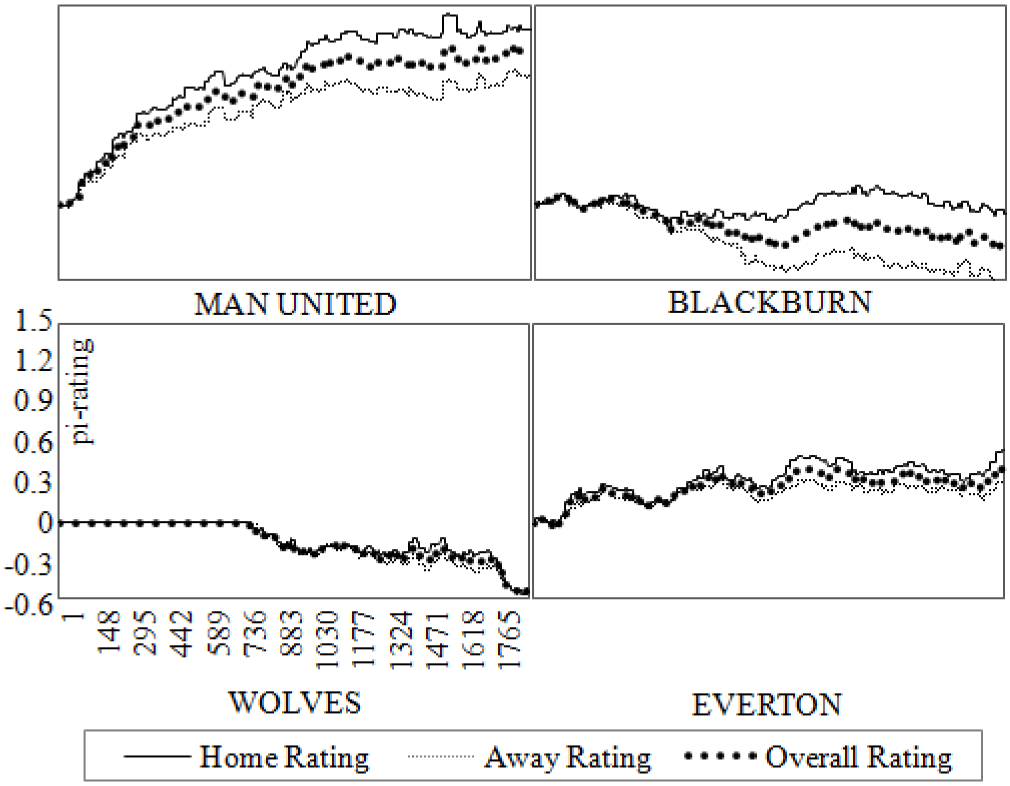


Figure 8.6. Development of the pi-ratings based on individual home and away performances for the specified teams[7] and from season 2007/08 to 2011/12 inclusive.

[7] For the newly promoted team Wolves the development of the ratings start at match instance 760 since no performances have been recorded relative to the residual EPL teams during the two preceding seasons.

8.4 Concluding remarks and future work

We have proposed a novel rating system, which we call pi-rating, for determining the level of ability of football teams on the basis of the relative discrepancies in scores through relevant match instances. The pi-rating is computationally efficient with minimal complexity and proceeds with dynamic modifications after every new match instance is observed by generating values that are meaningful in terms of diminished score difference relative to the average adversary within the league. The ratings can be used to formulate both score-based and result-based match predictions.

The rating system considers different ratings for when a team is playing at home and away, it also considers the relevant recent results to be more important than the former, and introduces diminished rewards for each additional goal difference greater than 1. Optimal learning rates ensure that the newly acquired match results are more important than the former and that the newly acquired information based on a home ground performance influences a team's ratings when playing away and vice versa, but also properly weighted for proceeding with appropriate rating modifications.

The pi-ratings were used as the basis of a football betting strategy against published market odds in an attempt to evaluate how well the rating values capture the ability of the various football teams. Over the period of the five most recent EPL seasons (2007/08 to 2011/12) the forecasting capability which was based on the generated rating values was sufficiently high to demonstrate profitability, even allowing for the bookmakers' build-in profit margin. This implies that the rating system properly captures the ability of a team at any time interval throughout the season.

This is the first academic study to demonstrate profitability against market odds using such a simple technique and hence, although the primary objective in this technique is rating and not prediction, the resulting ratings can be used as one of the model parameters for prediction purposes. In fact, the pi-ratings simplify the process for a forecasting football model in the sense that rating values reflect a team's current performance and thus, further factors and techniques that are normally introduced for determining the 'form' of a team by weighting the more recent results become redundant. Planned extensions of this research will determine:

a) the importance of the pi-ratings, by replacing relevant techniques of higher complexity for determining current team 'form', as inputs to the Bayesian network models that we have proposed in Chapters 6 and 7;

b) the value of pi-ratings in evaluating the relative ability of teams between different leagues by considering relevant match occurrences between teams of those leagues (e.g. Uefa Champions League). If successful, this will allow us to answer interesting questions such as *'which football league is best; the English Premier League or the Spanish La Liga?'*, and *'to what degree lower divisions differ from higher divisions in England'*, or even *'how much damage has the 2006 Italian football scandal, which was described as the biggest scandal in football history (Murali, 2011), caused to Serie A?'*.

CHAPTER 9

Concluding remarks and future directions

This chapter revisits the hypotheses of this research and evaluates the extent to which the results support those hypotheses. The chapter ends with potential future directions of this research project.

9.1 Revisiting the research hypotheses

The research hypotheses of this project are:

1. the Association Football gambling market publishes odds that are biased towards maximising profitability and hence, such odds suffer

from a degree of inaccuracy. This intended inefficiency, including any other that might be unknown, can be exploited using sophisticated probabilistic models that are sufficiently accurate for that matter.

2. Since the vast majority of the previous relevant academic studies (which have failed to demonstrate profitability that is consistent over time against published market odds) were solely focused on purely statistical and data-driven approaches to prediction, a novel BN model that considers both objective and subjective information for prediction (whereby subjective information represents information that is important for prediction but which historical data fails to capture) should be able to provide superior forecasting capability in an attempt to beat the market.

In (Dixon & Coles, 1997) the authors claimed 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, and (Graham & Stott, 2008) suggested that if such a work was particularly successful, it would not have been published.

9.2 Summary of results

The hypotheses are met to full extent. The most important results of this research are summarised, by Chapter, as follows:

- **Chapter 4**: Demonstrates that all of the various measures of accuracy used by all of the previous relevant academic studies for determining the forecast accuracy of football models are inadequate since they fail to recognise that football outcomes represent a ranked (ordinal) scale probability distribution. This raises severe concerns about the validity of conclusions from previous studies. We have proposed a well-established measure of accuracy, the Rank Probability Score (RPS), which has been missed by previous researcher, but which properly assesses football forecasting models. This work has been published by the Journal of Quantitative Analysis in Sports.

- **Chapter 5**: Provides numerous evidence of an (intended) inefficient Association Football gambling market. This work has been submitted for publication in an international academic journal.

- **Chapters 6 and 7:** A novel BN model was presented that was used to generate the EPL match forecasts during season 2010/11. This was the first academic study to demonstrate profitability against all of the (available) published market odds, and this work has been published by the Journal of Knowledge-Based Systems.

  A Bayesian network model that not only simplified the previously published model, but also provided improved forecasting capability by generating even higher profitability was used to generate the EPL match forecasts during season 2011/12. This work has been submitted for publication in an international academic journal.

Both of the models:

1. consider both objective and subjective information for prediction;
2. considered subjective indications by the same member of the research team, who is a football fan but definitely not an expert of the EPL;
3. demonstrated that subjective information improved the forecasting capability of the model significantly;
4. generated predictions before the matches were played, and predictions were published online at www.pi-football.com;
5. are easily applicable to any other football league;
6. emphasised the importance of Bayesian networks.

- **Chapter 8**: Presents a novel rating system (pi-rating) for determining the level of ability of football teams on the basis of the relative discrepancies in scores through relevant match instances. This rating system is computationally efficient with minimal complexity, and is the first academic study to demonstrate profitability against published market odds by using such a simple technique. This rating system proceeds with dynamic modifications after every new match instance is observed by generating values that are meaningful in terms of diminished score difference relative to the average adversary within that league. Furthermore, even though the models presented in Chapters 5 and 6 appeared to be particularly successful at beating bookmakers' odds, their forecasts did not incorporate score-based information about the relevant football teams. Therefore, the pi-ratings can be incorporated in football forecast models such as those of Chapters 5 and 6 in an attempt to further enhance their forecasting capability. In fact, the pi-ratings simplify the process for a forecasting football model in the sense that the rating values reflect a team's current performance and thus, further factors and techniques that are normally introduced for determining the 'form' of a football team by weighting the more recent results become redundant.

9.3 Possible future directions

The results of this Ph.D research suggest many possible future research directions. Below we enumerate some of them:

1. what appears to be missing from the academic literature is how bettors can take advantage of the various bonuses (e.g. deposit bonus) offered by many of the online bookmakers in an attempt to further increase profitability;

2. almost all of the past studies have only focused on $\{p(H), p(D), p(A)\}$ odds for deriving conclusions, primarily due to availability limitations. It would be very interesting to investigate how the betting markets behave for bets other than the standard football outcomes (i.e. players, goal-lines, cards, correct scores, tournament outrights etc.);

3. to investigate how the gambling market behaves during live betting. Live betting has emerged along with online betting and it has now

become exceptionally popular. In fact, bookmakers have reported that live betting accounts for the majority of the betting stakes (approximately 75% of the total volume of stakes has been reported by (bwin Group, 2010), which in turn represents a growth of approximately 7.1% from the previous year);

4. clearly the real potential benefits of our two models presented in Chapters 6 and 7 are critically dependent on both the structure of the model and the knowledge of the expert. A perfect BN model would still fail to beat the bookmakers at their own game if the subjective inputs are erroneous. Because of the weekly pressure to get all of the model predictions calculated and published online, there was inevitable inconsistency in the care and accuracy taken to consider all the subjective inputs for each match. In most cases the subjective inputs were provided by a member of the research team who is certainly not an expert on the English premier League. If the model were to be used by more informed experts we feel it would provide posterior beliefs of both higher precision and confidence;

5. to determine the importance of the pi-ratings as inputs to the Bayesian network models that we have proposed in Chapters 6 and 7;

6. to assess the value of pi-ratings in evaluating the relative ability of teams between different leagues by considering relevant match occurrences between teams of those leagues (e.g. Uefa Champions League). If successful, this will allow us to answer interesting questions such as *'which football league is best; the English Premier League or*

the Spanish La Liga?', and *'to what degree lower divisions differ from higher divisions in England'*, or even *'how much damage has the 2006 Italian football scandal, which was described as the biggest scandal in football history (Murali, 2011), caused to Serie A?'*;

7. to continue to provide football match forecasts online at www.pi-football.com and determine its value in terms of a potentially profitable business model. The models developed throughout this Ph.D research might serve as the basis for formulating even more powerful probabilistic football forecast models.

APPENDIX A

Evidence of an (intended) inefficient Association Football gambling market (Chapter 5)

Appendix A.1

The observations found for each competing team monitored were divided into the following six categories:

1. First team players missing the match (negative impact)
2. First team players returning back to action (positive impact)
3. Key player missing the match (negative impact)
4. Key player returning back to action (positive impact)
5. Managerial or ownership issues (positive or negative impact)
6. Other important factors (positive or negative impact)

We have introduced key-players as a distinct category[1] since each team normally has 1 or 2 key players which may cause significant problems to their teams if they are absent. For each observation, one of the two teams receives a score of +1 and the team with the highest score was expected to have the odds adjusted such that the probabilities for winning the particular match are increased in its favour. The results appear to be appealing and are summarised below:

- During the specified period of this study 252 matches were played. We have observed that 129 of those matches had their odds adjusted at least once by bwin, and 71 matches by Sportingbet, which translates to the respective adjustment rates of 64.50% and 35.50%. The resulting total of 200 cases considered 149 distinct matches; implying that

[1] Categories (3) and (4) are not subsets of (1) and (2).

59.12% of those 252 matches received at least one adjustment by at least one of the two bookmakers.

- In 98 out of the 200 cases only one bookmaker provided adjustments, whereas both bookmakers provided adjustments in the remaining 102 cases. However, 12 out of those 102 cases resulted in contradictory adjustment (e.g. bwin decreased the probability for a home win but Sportingbet increased the probability for that same outcome).
- In 63 cases the odds were adjusted on the day of the event, in 83 cases before the day of the event, and in 23 cases an adjustment was observed both before and on the day of the event[2].
- At least one cause was found (from categories 1 to 6 above) for each match adjustment in only 85 out of the 200 cases; implying that 57.5% of match adjustments could not have been explained by our selected factors.
- In 76 out of those 85 cases the evidence pointed towards one of the two competing teams. However, only in 47 out of those 76 cases (61.84%) does the evidence agree with the adjustment.

[2] We have missed this type of information for the first 31 observations and thus, we only report on a total of 169 observations.

Appendix A.2: Percentage shifts in published odds

Table A.2.1: Percentage shifts in published odds for bookmakers Sportingbet and bwin, from 07/11/2009 to 09/05/2010. A total of 200 occurrences are reported.

Match date	Bookmaker	Home team	Away team	Initial probabilities $p(H)$	$p(D)$	$p(A)$	Final probabilities $p(H)$	$p(D)$	$p(A)$	Central Tendency difference[3]
07/11/2009	bwin	Blackburn	Portsmouth	45.08	28.88	26.03	47.32	28.39	24.28	1.99
07/11/2009	bwin	Aston Villa	Bolton	54.30	25.64	20.07	52.77	27.16	20.07	0.77
07/11/2009	Sportingbet	Aston Villa	Bolton	54.98	25.92	19.10	52.67	27.56	19.77	1.49
07/11/2009	bwin	Man City	Burnley	68.55	20.56	10.89	72.25	18.50	9.25	2.67
07/11/2009	bwin	Tottenham	Sunderland	59.48	26.34	14.18	61.53	25.29	13.18	1.53
07/11/2009	Sportingbet	Wolves	Arsenal	12.07	21.55	66.39	11.32	21.57	67.11	0.73
08/11/2009	Sportingbet	Hull	Stoke	35.75	28.49	35.75	35.16	29.67	35.16	0.00
08/11/2009	bwin	Hull	Stoke	36.97	26.79	36.24	33.72	29.91	36.37	1.68
08/11/2009	Sportingbet	Chelsea	Man United	45.38	27.93	26.69	46.69	28.02	25.29	1.36
21/11/2009	Sportingbet	Liverpool	Man City	49.22	28.02	22.76	50.52	26.75	22.73	0.67
21/11/2009	bwin	Liverpool	Man City	49.89	28.40	21.72	49.84	27.12	23.05	0.69
21/11/2009	Sportingbet	Birmingham	Fulham	36.46	28.48	35.06	38.63	28.37	33.01	2.11
21/11/2009	bwin	Birmingham	Fulham	37.02	28.05	34.93	36.89	30.74	32.36	1.22
21/11/2009	Sportingbet	Burnley	Aston Villa	31.28	27.49	41.23	29.28	27.50	43.22	1.99
21/11/2009	bwin	Burnley	Aston Villa	32.36	27.53	40.10	29.36	27.61	43.02	2.96
21/11/2009	bwin	Chelsea	Wolves	76.92	15.38	7.69	76.86	14.76	8.38	0.38
21/11/2009	bwin	Hull	West Ham	34.24	28.02	37.74	33.57	27.97	38.46	0.70
21/11/2009	Sportingbet	Sunderland	Arsenal	15.15	24.24	60.61	17.24	25.14	57.62	2.54
21/11/2009	bwin	Sunderland	Arsenal	14.77	25.65	59.58	16.03	26.34	57.62	1.61
22/11/2009	Sportingbet	Bolton	Blackburn	43.43	28.06	28.50	39.47	29.76	30.77	3.12
22/11/2009	bwin	Bolton	Blackburn	52.74	26.75	20.51	41.90	28.36	29.74	10.03
22/11/2009	Sportingbet	Tottenham	Wigan	64.80	22.68	12.51	62.74	24.26	13.00	1.27
22/11/2009	Sportingbet	Stoke	Portsmouth	43.22	27.50	29.28	44.22	27.89	27.89	1.19
28/11/2009	bwin	Man City	Hull	71.24	19.50	9.26	73.85	18.46	7.69	2.09
28/11/2009	bwin	Aston Villa	Tottenham	40.10	27.53	32.36	39.36	27.61	33.03	0.71
29/11/2009	Sportingbet	Arsenal	Chelsea	36.46	28.48	35.06	35.59	28.81	35.59	0.70
29/11/2009	Sportingbet	Everton	Liverpool	31.81	27.90	40.29	27.46	28.32	44.21	4.13
29/11/2009	bwin	Everton	Liverpool	29.86	28.05	42.08	28.00	28.00	44.00	1.89
05/12/2009	Sportingbet	Arsenal	Stoke	72.73	18.18	9.09	74.38	16.53	9.09	0.83
05/12/2009	Sportingbet	Blackburn	Liverpool	21.57	26.65	51.78	20.12	26.63	53.25	1.47
05/12/2009	bwin	Blackburn	Liverpool	20.59	26.47	52.94	18.45	25.63	55.92	2.56
05/12/2009	Sportingbet	West Ham	Man United	14.55	22.73	62.72	12.15	22.78	65.08	2.38
05/12/2009	bwin	West Ham	Man United	12.86	24.65	62.49	12.33	22.55	65.12	1.58
05/12/2009	bwin	Man City	Chelsea	23.36	26.75	49.89	21.57	26.89	51.54	1.72
06/12/2009	Sportingbet	Everton	Tottenham	35.06	28.48	36.46	32.34	28.30	39.37	2.81
06/12/2009	bwin	Everton	Tottenham	35.51	27.56	36.93	29.86	28.05	42.08	5.40
12/12/2009	bwin	Bolton	Man City	23.05	27.12	49.84	21.04	26.07	52.89	2.53
12/12/2009	bwin	Chelsea	Everton	73.85	18.46	7.69	75.79	16.81	7.40	1.12
12/12/2009	Sportingbet	Man united	Aston Villa	67.04	20.89	12.07	62.81	22.68	14.51	3.34
12/12/2009	bwin	Man united	Aston Villa	65.96	21.73	12.31	65.06	21.74	13.20	0.89

[3] Difference between ordinal distribution means with values $\{0, 0.5, 1\}$ for outcomes $\{p(H), p(D), p(A)\}$ respectively.

13/12/2009	Sportingbet	Liverpool	Arsenal	41.31	28.40	30.29	43.24	28.38	28.38	1.92
13/12/2009	bwin	Liverpool	Arsenal	43.04	28.92	28.04	45.12	28.46	26.43	1.85
16/12/2009	Sportingbet	Burnley	Arsenal	15.15	24.24	60.61	14.03	22.80	63.17	1.84
16/12/2009	bwin	Burnley	Arsenal	14.19	22.22	63.59	13.17	20.96	65.87	1.64
16/12/2009	bwin	Chelsea	Portsmouth	80.23	13.18	6.59	82.62	11.94	5.44	1.77
16/12/2009	Sportingbet	Liverpool	Wigan	69.77	20.16	10.08	72.63	19.11	8.25	2.34
16/12/2009	Sportingbet	Tottenham	Man City	40.34	28.36	31.30	42.28	29.32	28.40	2.42
16/12/2009	bwin	Tottenham	Man City	41.12	28.04	30.84	42.08	28.05	29.86	0.97
09/01/2010	Sportingbet	Arsenal	Everton	66.39	21.55	12.07	68.22	21.10	10.67	1.61
09/01/2010	bwin	Arsenal	Everton	66.12	22.31	11.57	68.48	21.25	10.27	1.83
09/01/2010	Sportingbet	Birmingham	Man United	15.15	24.24	60.61	15.74	25.86	58.40	1.40
09/01/2010	bwin	Birmingham	Man United	16.82	25.35	57.83	15.40	24.98	59.62	1.61
16/01/2010	Sportingbet	Stoke	Liverpool	15.75	25.16	59.09	26.63	29.21	44.17	12.90
16/01/2010	bwin	Stoke	Liverpool	13.70	24.66	61.64	24.96	28.86	46.18	13.36
16/01/2010	bwin	Wolves	Wigan	40.20	28.45	31.35	39.33	29.34	31.33	0.43
16/01/2010	bwin	Everton	Man City	31.35	28.45	40.20	34.20	28.86	36.94	3.06
17/01/2010	Sportingbet	Blackburn	Fulham	41.31	28.40	30.29	40.23	28.29	31.48	1.13
17/01/2010	bwin	Blackburn	Fulham	43.04	28.92	28.04	40.98	29.27	29.75	1.88
17/01/2010	Sportingbet	Bolton	Arsenal	15.75	25.16	59.09	15.15	22.17	62.68	2.10
17/01/2010	bwin	Bolton	Arsenal	14.77	25.65	59.58	14.20	24.28	61.52	1.26
20/01/2010	Sportingbet	Arsenal	Bolton	75.86	16.55	7.59	76.59	16.45	6.96	0.68
20/01/2010	bwin	Arsenal	Bolton	73.85	18.46	7.69	78.41	15.42	6.17	3.04
20/01/2010	bwin	Liverpool	Tottenham	40.16	27.99	31.85	41.12	28.04	30.84	0.99
26/01/2010	Sportingbet	Portsmouth	West Ham	38.63	28.37	33.01	40.34	28.36	31.30	1.71
26/01/2010	bwin	Portsmouth	West Ham	40.20	28.45	31.35	39.47	28.54	31.99	0.69
26/01/2010	Sportingbet	Wolves	Liverpool	20.12	26.63	53.25	17.26	26.65	56.10	2.85
26/01/2010	bwin	Wolves	Liverpool	20.59	26.47	52.94	15.79	24.63	59.58	5.72
26/01/2010	Sportingbet	Bolton	Burnley	50.52	26.75	22.73	53.25	26.63	20.12	2.68
26/01/2010	bwin	Bolton	Burnley	51.46	27.24	21.29	54.41	25.69	19.89	2.18
27/01/2010	Sportingbet	Aston Villa	Arsenal	32.34	28.30	39.37	29.32	28.40	42.28	2.96
27/01/2010	bwin	Aston Villa	Arsenal	30.36	28.49	41.15	28.00	28.00	44.00	2.60
27/01/2010	bwin	Chelsea	Birmingham	73.98	17.61	8.41	75.58	16.04	8.38	0.81
27/01/2010	bwin	Blackburn	Wigan	47.47	28.48	24.04	48.62	27.99	23.39	0.90
27/01/2010	Sportingbet	Everton	Sunderland	54.98	25.92	19.10	56.66	25.90	17.43	1.67
27/01/2010	bwin	Everton	Sunderland	54.44	26.07	19.49	57.76	24.64	17.60	2.60
30/01/2010	Sportingbet	Liverpool	Bolton	63.17	22.80	14.03	65.08	22.78	12.15	1.90
30/01/2010	bwin	Liverpool	Bolton	65.01	23.08	11.91	65.94	22.52	11.54	0.65
30/01/2010	Sportingbet	West Ham	Blackburn	43.24	28.38	28.38	32.34	28.30	39.37	10.95
31/01/2010	bwin	Man City	Portsmouth	71.24	19.50	9.26	73.85	18.46	7.69	2.09
31/01/2010	bwin	Arsenal	Man United	38.46	27.97	33.57	37.74	28.02	34.24	0.70
01/02/2010	bwin	Sunderland	Stoke	47.37	27.99	24.63	46.23	28.45	25.33	0.92
03/02/2010	bwin	Fulham	Portsmouth	52.88	28.04	19.08	55.97	26.77	17.26	2.45
06/02/2010	bwin	Liverpool	Everton	52.88	27.22	19.90	54.24	26.35	19.41	0.92
06/02/2010	bwin	Burnley	West Ham	39.25	28.38	32.37	37.70	29.32	32.98	1.09
06/02/2010	Sportingbet	Bolton	Fulham	40.34	28.36	31.30	43.24	28.38	28.38	2.91
06/02/2010	bwin	Sunderland	Wigan	43.99	28.43	27.58	46.23	28.45	25.33	2.24
07/02/2010	bwin	Chelsea	Arsenal	52.74	26.75	20.51	54.24	26.35	19.41	1.30
09/02/2010	bwin	Man City	Bolton	68.50	21.76	9.73	65.01	23.08	11.91	2.84
09/02/2010	bwin	Portsmouth	Sunderland	37.74	28.02	34.24	38.46	27.97	33.57	0.70
09/02/2010	bwin	Wigan	Stoke	43.99	28.43	27.58	43.05	28.48	28.48	0.92
09/02/2010	bwin	Fulham	Burnley	57.63	25.61	16.76	56.00	26.40	17.60	1.23
10/02/2010	bwin	Aston Villa	Man United	25.97	27.94	46.09	22.06	26.47	51.47	4.64
10/02/2010	bwin	Blackburn	Hull	52.74	26.75	20.51	51.33	27.18	21.49	1.19
16/02/2010	Sportingbet	Stoke	Man City	26.69	27.93	45.38	35.73	32.38	31.88	11.27

16/02/2010	bwin	Stoke	Man City	26.43	28.46	45.12	29.74	28.36	41.90	3.26
17/02/2010	bwin	Wigan	Bolton	41.17	29.88	28.95	42.11	28.95	28.95	0.47
20/02/2010	bwin	Everton	Man United	18.48	27.17	54.35	19.41	26.35	54.24	0.52
20/02/2010	Sportingbet	West Ham	Hull	50.66	27.63	21.71	52.44	27.44	20.12	1.68
20/02/2010	bwin	West Ham	Hull	52.94	26.47	20.59	54.24	26.35	19.41	1.24
20/02/2010	Sportingbet	Arsenal	Sunderland	76.71	15.74	7.54	76.59	16.45	6.96	0.23
20/02/2010	bwin	Portsmouth	Stoke	38.53	28.45	33.02	40.20	28.45	31.35	1.68
21/02/2010	bwin	Aston Villa	Burnley	65.01	23.08	11.91	68.55	20.56	10.89	2.28
21/02/2010	bwin	Fulham	Birmingham	43.05	28.48	28.48	45.12	28.46	26.43	2.06
21/02/2010	bwin	Man City	Liverpool	36.94	28.86	34.20	37.64	28.82	33.54	0.69
27/02/2010	bwin	Chelsea	Man City	63.89	23.75	12.35	65.94	22.52	11.54	1.43
27/02/2010	Sportingbet	Birmingham	Wigan	44.22	27.89	27.89	48.02	27.65	24.33	3.68
27/02/2010	bwin	Birmingham	Wigan	47.37	27.99	24.63	48.62	27.99	23.39	1.25
27/02/2010	bwin	Bolton	Wolves	50.10	28.09	21.81	48.62	27.99	23.39	1.53
27/02/2010	Sportingbet	Burnley	Portsmouth	40.29	27.90	31.81	45.38	27.93	26.69	5.10
27/02/2010	bwin	Burnley	Portsmouth	44.00	28.00	28.00	48.68	27.61	23.71	4.48
28/02/2010	Sportingbet	Tottenham	Everton	47.82	27.95	24.23	45.53	28.46	26.02	2.04
28/02/2010	bwin	Tottenham	Everton	47.37	27.99	24.63	43.98	28.86	27.16	2.96
28/02/2010	Sportingbet	Sunderland	Fulham	41.28	27.94	30.78	45.38	27.93	26.69	4.10
06/03/2010	Sportingbet	Arsenal	Burnley	79.88	14.03	6.08	81.51	13.10	5.39	1.16
08/03/2010	Sportingbet	Wigan	Liverpool	18.32	26.17	55.51	18.13	24.17	57.70	1.19
09/03/2010	bwin	Portsmouth	Birmingham	35.56	28.89	35.56	36.94	28.86	34.20	1.37
13/03/2010	Sportingbet	Tottenham	Blackburn	62.74	24.26	13.00	64.87	22.15	12.97	1.08
13/03/2010	Sportingbet	Birmingham	Everton	33.68	28.42	37.89	32.34	28.30	39.37	1.41
13/03/2010	bwin	Birmingham	Everton	33.54	28.82	37.64	31.96	29.42	38.62	1.28
13/03/2010	Sportingbet	Bolton	Wigan	43.24	28.38	28.38	44.21	28.32	27.46	0.94
13/03/2010	bwin	Chelsea	West Ham	75.56	16.76	7.68	76.92	15.38	7.69	0.68
13/03/2010	bwin	Stoke	Aston Villa	30.36	28.49	41.15	29.74	28.36	41.90	0.69
13/03/2010	Sportingbet	Hull	Arsenal	13.46	21.63	64.90	13.42	20.13	66.45	0.80
13/03/2010	bwin	Hull	Arsenal	11.56	24.66	63.78	12.31	21.73	65.96	0.71
14/03/2010	Sportingbet	Man United	Fulham	74.41	16.51	9.08	75.88	15.84	8.28	1.14
14/03/2010	bwin	Man United	Fulham	73.98	17.61	8.41	75.58	16.04	8.38	0.81
14/03/2010	Sportingbet	Sunderland	Man City	26.69	27.93	45.38	27.89	27.89	44.22	1.18
16/03/2010	bwin	Wigan	Aston Villa	28.00	28.00	44.00	25.70	28.04	46.26	2.28
20/03/2010	Sportingbet	Stoke	Tottenham	27.46	28.32	44.21	28.38	28.38	43.24	0.94
20/03/2010	bwin	Stoke	Tottenham	29.40	28.50	42.10	27.58	28.43	43.99	1.86
20/03/2010	bwin	Sunderland	Birmingham	42.08	28.05	29.86	45.08	28.88	26.03	3.42
21/03/2010	bwin	Man united	Liverpool	55.92	25.63	18.45	57.63	25.61	16.76	1.70
21/03/2010	bwin	Fulham	Man City	33.54	28.82	37.64	31.35	28.45	40.20	2.38
24/03/2010	bwin	Aston Villa	Sunderland	60.30	23.66	16.04	60.30	23.66	16.04	0.00
24/03/2010	Sportingbet	Man City	Everton	50.52	26.75	22.73	47.66	27.86	24.47	2.30
24/03/2010	bwin	Man City	Everton	51.28	25.64	23.08	50.50	26.40	23.10	0.41
24/03/2010	bwin	Portsmouth	Chelsea	9.26	19.50	71.24	9.24	16.81	73.95	1.36
27/03/2010	bwin	Chelsea	Aston Villa	66.17	21.05	12.78	67.04	21.03	11.94	0.85
27/03/2010	Sportingbet	West Ham	Stoke	45.38	27.93	26.69	44.22	27.89	27.89	1.18
27/03/2010	bwin	West Ham	Stoke	46.26	28.04	25.70	45.12	28.46	26.43	0.93
28/03/2010	Sportingbet	Burnley	Blackburn	36.46	28.48	35.06	35.75	28.49	35.75	0.70
28/03/2010	bwin	Burnley	Blackburn	35.56	28.89	35.56	36.24	28.88	34.88	0.68
03/04/2010	bwin	Man united	Chelsea	43.99	28.43	27.58	36.94	28.86	34.20	6.84
03/04/2010	bwin	Bolton	Aston Villa	29.86	28.05	42.08	32.51	28.07	39.42	2.65
03/04/2010	bwin	Portsmouth	Blackburn	33.02	28.45	38.53	28.00	28.00	44.00	5.25
03/04/2010	bwin	Stoke	Hull	54.34	28.87	16.80	52.77	27.99	19.24	2.00
03/04/2010	bwin	Sunderland	Tottenham	27.58	28.43	43.99	27.33	27.33	45.33	0.79
04/04/2010	bwin	Everton	West Ham	63.71	23.09	13.20	66.17	21.05	12.78	1.44

04/04/2010	bwin	Birmingham	Liverpool	20.07	27.16	52.77	20.98	27.15	51.87	0.90
04/04/2010	Sportingbet	Fulham	Wigan	47.82	27.95	24.23	43.24	27.94	28.82	4.59
04/04/2010	bwin	Fulham	Wigan	47.53	28.08	24.39	44.00	28.00	28.00	3.57
09/04/2010	bwin	Blackburn	Man United	13.70	24.66	61.64	13.36	21.70	64.94	1.81
11/04/2010	Sportingbet	Liverpool	Wigan	69.77	20.16	10.08	72.73	18.18	9.09	1.97
11/04/2010	Sportingbet	Man City	Birmingham	65.08	22.78	12.15	66.39	21.55	12.07	0.70
13/04/2010	bwin	Chelsea	Bolton	80.19	13.66	6.15	82.62	11.22	6.17	1.20
14/04/2010	Sportingbet	Aston Villa	Everton	41.31	28.40	30.29	40.34	28.36	31.30	0.99
14/04/2010	bwin	Aston Villa	Everton	46.26	28.04	25.70	41.12	28.04	30.84	5.14
14/04/2010	bwin	Wigan	Portsmouth	57.76	24.64	17.60	60.30	23.66	16.04	2.05
14/04/2010	bwin	Tottenham	Arsenal	31.85	27.99	40.16	30.81	27.18	42.01	1.45
17/04/2010	Sportingbet	Man City	Man United	36.26	27.47	36.26	34.83	27.44	37.73	1.45
17/04/2010	Sportingbet	Birmingham	Hull	52.17	26.09	21.74	52.50	25.91	21.59	0.24
17/04/2010	bwin	Fulham	Wolves	50.50	26.40	23.10	45.12	28.46	26.43	4.35
17/04/2010	bwin	Stoke	Bolton	49.97	28.02	22.01	47.32	28.39	24.28	2.46
18/04/2010	Sportingbet	Portsmouth	Aston Villa	19.10	25.92	54.98	18.13	24.17	57.70	1.84
18/04/2010	bwin	Portsmouth	Aston Villa	19.50	25.03	55.47	16.78	23.67	59.55	3.40
19/04/2010	bwin	Liverpool	West Ham	67.04	21.03	11.94	68.39	20.07	11.54	0.87
21/04/2010	bwin	Hull	Aston Villa	30.36	28.49	41.15	26.82	28.04	45.14	3.76
24/04/2010	bwin	Man United	Tottenham	63.93	21.81	14.26	64.20	21.01	14.79	0.13
24/04/2010	bwin	Bolton	Portsmouth	59.55	23.67	16.78	61.47	22.49	16.04	1.33
24/04/2010	bwin	West Ham	Wigan	52.88	27.22	19.90	51.57	25.79	22.64	2.02
24/04/2010	Sportingbet	Wolves	Blackburn	40.34	28.36	31.30	44.22	27.89	27.89	3.64
24/04/2010	bwin	Wolves	Blackburn	39.35	29.83	30.82	42.08	29.86	28.05	2.75
24/04/2010	bwin	Arsenal	Man City	49.06	26.35	24.59	47.29	25.98	26.73	1.95
25/04/2010	Sportingbet	Aston Villa	Birmingham	56.68	25.19	18.14	57.62	25.14	17.24	0.92
25/04/2010	bwin	Aston Villa	Birmingham	57.76	24.64	17.60	59.55	23.67	16.78	1.31
01/05/2010	bwin	Portsmouth	Wolves	33.00	27.18	39.83	40.23	28.92	30.85	8.11
01/05/2010	bwin	Stoke	Everton	26.43	28.46	45.12	26.35	27.53	46.12	0.54
01/05/2010	bwin	Tottenham	Bolton	73.98	17.61	8.41	73.95	16.81	9.24	0.43
02/05/2010	Sportingbet	Liverpool	Chelsea	25.29	26.78	47.92	23.83	25.87	50.30	1.92
02/05/2010	bwin	Liverpool	Chelsea	33.09	25.74	41.18	22.54	24.65	52.81	11.09
02/05/2010	bwin	Fulham	West Ham	41.08	27.59	31.33	38.58	28.93	32.49	1.83
02/05/2010	Sportingbet	Sunderland	Man United	13.04	21.74	65.22	12.07	21.55	66.39	1.08
02/05/2010	bwin	Sunderland	Man United	12.31	21.73	65.96	10.84	18.25	70.90	3.21
03/05/2010	Sportingbet	Wigan	Hull	50.44	27.94	21.62	53.28	25.16	21.56	1.44
03/05/2010	bwin	Wigan	Hull	52.09	26.49	21.41	53.76	25.69	20.55	1.27
03/05/2010	bwin	Blackburn	Arsenal	28.46	26.43	45.12	27.24	26.46	46.30	1.20
05/05/2010	bwin	Man City	Tottenham	50.00	25.00	25.00	48.63	26.40	24.97	0.67
09/05/2010	Sportingbet	Arsenal	Fulham	75.86	16.55	7.59	79.03	13.98	6.99	1.88
09/05/2010	bwin	Arsenal	Fulham	78.37	14.23	7.40	79.46	13.17	7.37	0.55
09/05/2010	bwin	Bolton	Birmingham	42.08	28.05	29.86	44.00	28.00	28.00	1.89
09/05/2010	Sportingbet	Burnley	Tottenham	10.65	19.69	69.66	15.79	21.61	62.60	6.10
09/05/2010	bwin	Burnley	Tottenham	14.41	21.96	63.62	13.70	22.55	63.76	0.43
09/05/2010	bwin	Chelsea	Wigan	84.00	10.56	5.44	85.43	9.71	4.86	1.00
09/05/2010	Sportingbet	Everton	Portsmouth	71.01	18.25	10.74	74.17	15.76	10.07	1.91
09/05/2010	bwin	Everton	Portsmouth	71.20	18.51	10.28	73.95	16.81	9.24	1.89
09/05/2010	bwin	Hull	Liverpool	16.07	24.32	59.61	15.76	21.95	62.29	1.49
09/05/2010	bwin	Man United	Stoke	82.62	11.94	5.44	83.97	10.26	5.77	0.51
09/05/2010	Sportingbet	West Ham	Man City	19.10	24.19	56.71	25.16	25.88	48.96	6.90
09/05/2010	bwin	West Ham	Man City	23.18	24.73	52.09	24.33	25.69	49.98	1.63
09/05/2010	bwin	Wolves	Sunderland	40.23	28.92	30.85	41.12	28.04	30.84	0.45

APPENDIX B

pi-football: A Bayesian network model for forecasting Association Football match outcomes (Chapter 6)

Appendix B.1: Subjective scenarios and assumptions per specified variable (node)

Table B.1.1. Team Strength (as presented in Figure 6.2)

ID	Variable (node)	Description	Subjective Scenarios
I.	Subjective team strength (in points)	Expert indication regarding the current strength of the team in seasonal points.	[0,114]
II.	Confidence	Expert indication regarding its confidence about his input (I).	[Very High, High, Medium, Low, Very Low]
III.	Current Points	Assumption: Variance as demonstrated in Figure 6.1, given variable "Number of matches played".	-
IV.	Points during season 2005/06	Assumption: variance=(Variance+3^6)	-
V.	Points during season 2006/07	Assumption: variance=(Variance+3^5)	-
VI.	Points during season 2007/08	Assumption: variance=(Variance+3^4)	-
VII.	Points during season 2008/09	Assumption: variance=(Variance+3^3)	-
VIII.	Points during season 2009/10	Assumption: variance=(Variance+3^2)	-
IX.	Predicted mean (in points)	The predicted team strength after considering all of the seven parameters Assumption: mean=57, variance=300	-

Table B.1.2. Team Form (as presented in Figure 6.3)

ID	Variable (node)	Description	Subjective Scenarios
I.	Primary key-player availability	Expert indication regarding his confidence about the availability of the primary key-player.	[Very High, High, Medium, Low, Very Low]
II.	Secondary key-player availability	Expert indication regarding his confidence about the availability of the secondary key-player.	[Very High, High, Medium, Low, Very Low]
III.	Tertiary key-player availability	Expert indication regarding his confidence about the availability of the tertiary key-player.	[Very High, High, Medium, Low, Very Low]
IV.	Remaining first team players availability	Expert indication regarding his confidence about the availability of the remaining first-team players.	[Very High, High, Medium, Low, Very Low]
V.	First team players returning	Expert indication regarding the potential return of other first team players who missed the last few matches.	[Very High, High, Medium, Low, Very Low]

Table B.1.3. Team Psychology (as presented in Figure 6.4)

ID	Variable (node)	Description	Subjective Scenarios
I.	Team spirit and motivation	Expert indication regarding the team's level of motivation and team spirit	[Very High, High, Normal, Low, Very Low]
II.	Confidence	Expert indication regarding its confidence about his input in (I).	[Very High, High, Medium, Low, Very Low]
III.	Managerial impact	Expert indication regarding the impact of the current managerial situation.	[Very High, High, Normal, Low, Very Low]
IV.	Head-to-Head bias	Expert indication regarding potential biases in a head-to-head encounter between the two teams.	[High advantage for home team, Advantage for home team, No bias, Advantage for away team, High advantage for away team]

Table B.1.4. Team Fatigue (as presented in Figure 6.5)

ID	Variable (node)	Description	Subjective Scenarios
I.	Toughness of previous match	Expert indication regarding the toughness of previous match.	[Lowest, Very Low, Low, Medium, High, Very High, Highest]
II.	First team players rested during last match	Expert indication regarding the first team players rested during last match.	[1-2, 3, 4, 5, 6+]
III.	National team participation	Expert indication regarding the level of international participation by the first team players.	[None, Few, Half team, Many, All]

Appendix B.2: An actual example of component's 1 process (as presented in Fig. 6.2)

Figure B2.1 presents a real component 1 example between Manchester City (home team) and Manchester United, as prepared for the 11th of October 2010. The steps for calculating component's 1 forecast are enumerated below:

1) **Previous information**: the points accumulated per previous season are passed as five distinct ordered inputs. Starting from the oldest season, the inputs are $[43, 42, 55, 50, 67]$ for Man City, and $[83, 89, 87, 90, 85]$ for Man United. Note that Man City generates a significantly higher variance than that of Man United, with the more recent seasons having greater impact as described and illustrated in Section 6.3.1.

2) **Current information:** the points accumulated for the current season, as well as the total number of matches played are passed as a single parameter with the appropriate variance as described and illustrated in Section 6.3.1. For Man City the inputs are $[20, 11]$ and for Man United the inputs are $[23, 11]$, for points accumulated and number of matches played respectively.

3) **Subjective information** (optional): the optional subjective indication about the current team's strength in total points, as well as the confidence with reference to that indication are passed as a single parameter. For Man City, we suggested that the team was playing as a 72-point team (a 5-point increase from last season) with *High*

confidence (out of *Very High*)[1]. On the other hand, we have introduced a 5-point decrease for Man United with *High* confidence[2]. Accordingly, the inputs were $[72, High]$ and $[80, High]$ for Man City and Man United respectively.

4) The model summarises the seven parameters in node *Mean*. The impact each parameter has is dependent on its certainty (variance). For Man City the summarised belief in total points (node *Mean*) is 68.95 whereas for Man United is 80.78. Note that the variance introduced for Man City is a higher than that of Man United; 26.83 and 21.92 respectively.

5) Each team's *Mean* is converted in the predetermined 14-scale ranking. The model suggests that Man City will most likely perform similar to teams ranked 3 to 4 (out of 14), whereas for Man United it mostly suggests ranks 1 and 2.

6) The model generates the objective forecast in node *Match Prediction*, by considering each teams estimated ranking, before proceeding to potential forecast revisions suggested by the expert constructed component models 2, 3 and 4.

[1] A 5-point increase was suggested due to high profile players joining the team during the summer transfer window.

[2] A 5-point decrease was suggested due to the significant decrease in stamina observed by the older core-team players (e.g. Scholes, Giggs, Ferdinand, Vidic) without taking care of appropriate replacements.

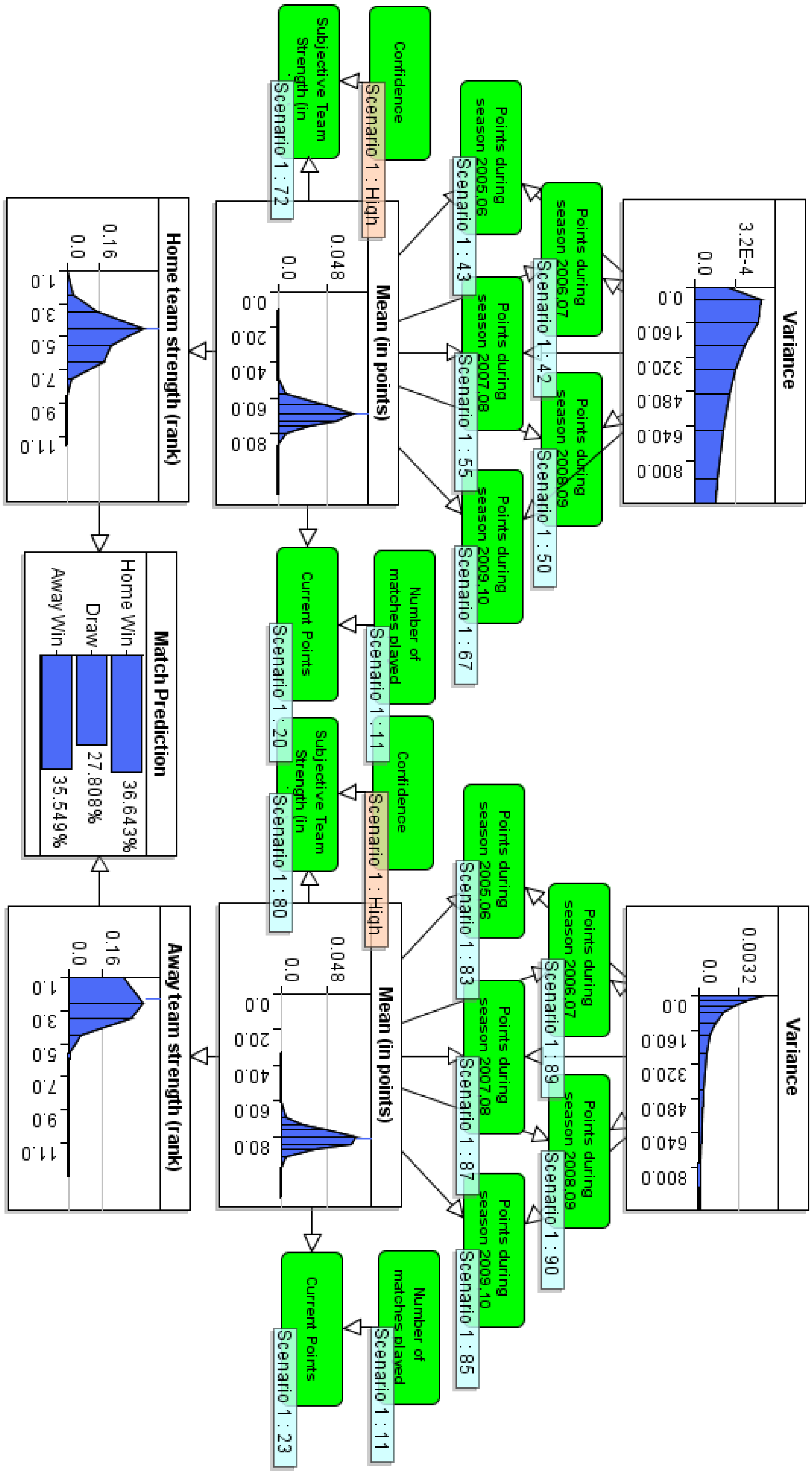


Figure B.2.1. An actual example of the Bayesian network (from Figure 6.3) at component 1. The parameters represent the actual observations provided from the Man City vs. Man United match, 10th of November, 2010.

Appendix B.3: Match RPS per dataset

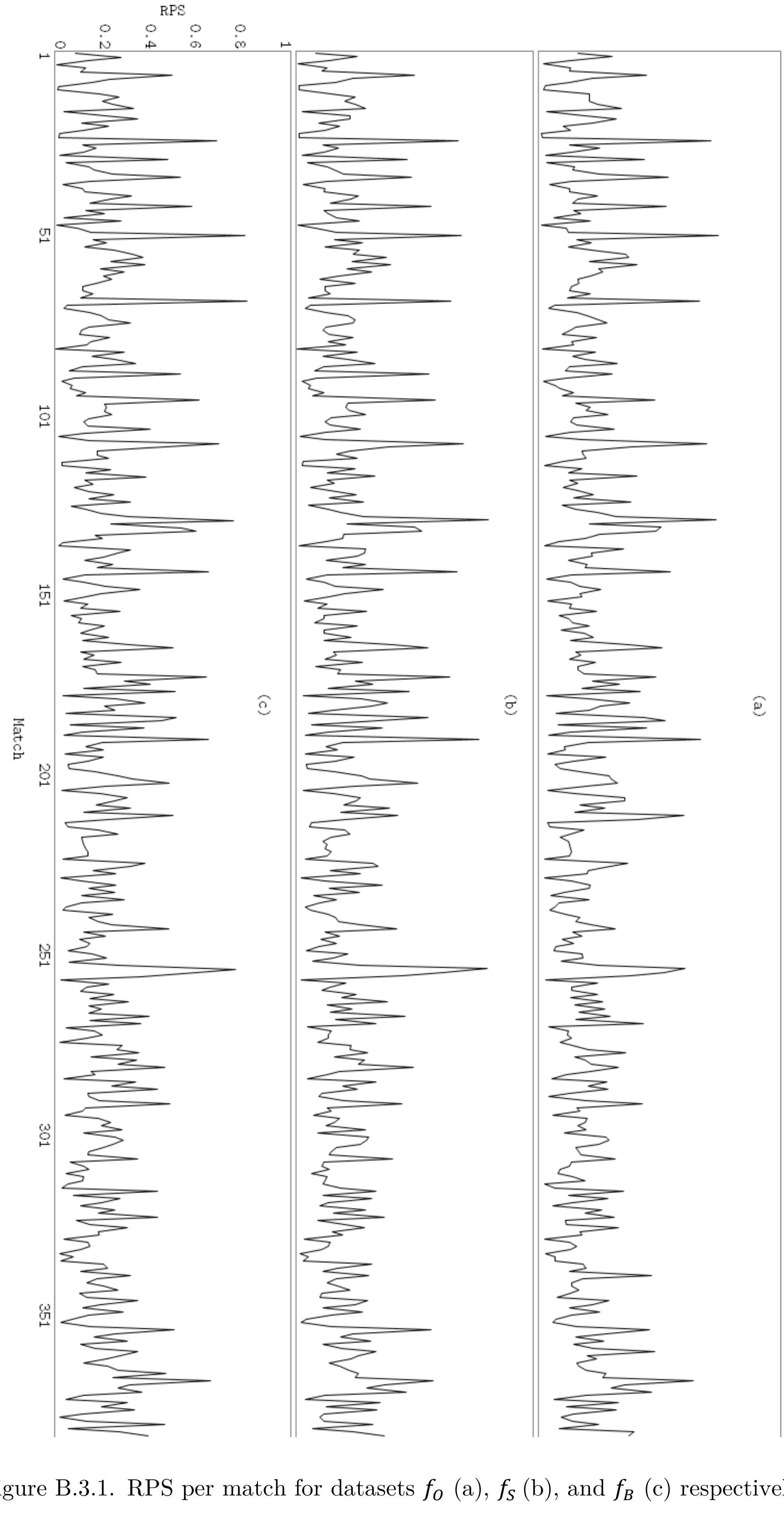


Figure B.3.1. RPS per match for datasets f_O (a), f_S (b), and f_B (c) respectively.

Appendix B.4: Evidence of significant improvements in f_O by subjective information

In this section we provide evidence of football matches in which subjective information revised f_O the most. Table B.4.1 presents 17 with the highest absolute RPS discrepancies between f_O and f_S forecasts, assuming a minimum discrepancy level of 0.1. The instances are ranked by highest discrepancy and the 'Decision' column indicates whether the subjective information improved f_O.

Overall, the results appear to be particularly encouraging. Only in 6 out of the 17 cases our subjective information leads to a higher forecast error. The results are even more encouraging when we only concentrate on the first 10 highest discrepancy instances, in which subjective revisions improve 8 out of the 10 instances. Further, in those 17 instances we have observed 15 distinct teams, and no evidence exist that strong subjective indications follow a particular type of a team. A rather surprising and interesting observation is that the observed outcome is a draw in only in 1 out of the 17 instances presented here.

Table B.4.1: RPS discrepancies ≥ 0.1 between objective (f_O) and revised (f_S); ranked by highest discrepancy

RPS Discrep.	Date	Home Team	Away Team	R	Objective (f_O) p(H)	Objective (f_O) p(D)	Objective (f_O) p(A)	Revised (f_S) p(H)	Revised (f_S) p(D)	Revised (f_S) p(A)	Decision
.2078	14/05/2011	Sunderland	Wolves	A	.4942	.3403	.1656	.2627	.4124	.3250	✓
.1765	06/03/2011	Liverpool	Man Utd	H	.2392	.2219	.5389	.3423	.3691	.2887	✓
.1614	03/10/2010	Liverpool	Blackpool	A	.8303	.1412	.0285	.6516	.2895	.0589	✓
.1582	09/04/2011	Man Utd	Fulham	H	.7570	.1881	.0549	.4016	.4552	.1432	✗
.1421	22/05/2011	Stoke	Wigan	A	.5140	.3023	.1837	.3535	.3684	.2781	✓
.1406	02/10/2010	Sunderland	Man Utd	D	.1223	.1940	.6837	.2029	.3973	.3998	✓
.1322	18/09/2010	Tottenham	Wolves	H	.7422	.1751	.0827	.4396	.4063	.1541	✗
.1307	06/11/2010	Bolton	Tottenham	H	.2519	.2523	.4958	.3384	.3358	.3259	✓
.1270	22/08/2010	Newcastle	Aston Villa	H	.2693	.3161	.4146	.3828	.3514	.2658	✓
.1228	25/01/2011	Wigan	Aston Villa	A	.3436	.3431	.3133	.2058	.3433	.4508	✓
.1219	29/12/2010	Liverpool	Wolves	A	.7162	.1717	.1121	.8058	.1406	.0536	✗
.1156	23/04/2011	Sunderland	Wigan	H	.4138	.3310	.2552	.2848	.3568	.3584	✗
.1150	01/02/2011	Sunderland	Chelsea	A	.2661	.3861	.3478	.1556	.3363	.5082	✓
.1104	27/12/2010	Arsenal	Chelsea	H	.4034	.3383	.2583	.2828	.3578	.3594	✗
.1102	28/12/2010	Sunderland	Blackpool	A	.5200	.2791	.2009	.3929	.3380	.2692	✓
.1063	25/09/2010	Arsenal	West Br.	A	.8196	.1499	.0305	.7063	.2424	.0512	✓
.1023	22/01/2011	Wolves	Liverpool	A	.3070	.3465	.3466	.4038	.3465	.2497	✗

Appendix B.5: Betting simulation given objective forecasts

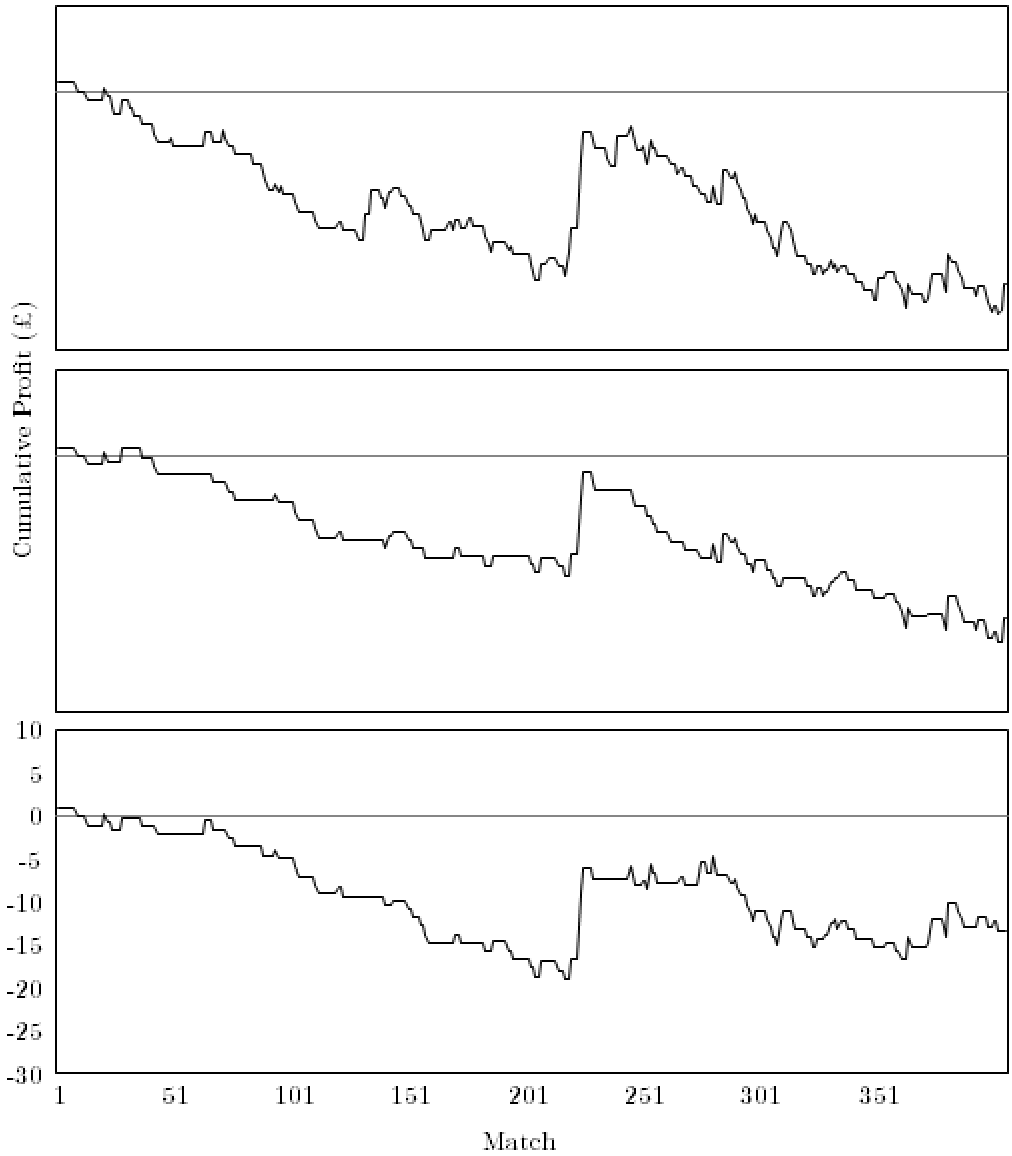


Figure B.5.1. Cumulative profit/loss observed given f_O when simulating the standard betting strategy at discrepancy levels of ≥ 5% against a) f_{maxB}, b) f_{meanB} and c) f_{WH}.

Appendix B.6: Betting simulation at different levels of discrepancy given f_S

Table B.6.1. Betting simulation stats given f_S against) f_{maxB}, b) f_{meanB} and c) f_{WH} at discrepancy levels from 1% to 20%

	Maximum odds			Mean odds			William Hill odds		
Discrepancy	No. of bets	Returns (£)	Profit/Loss (£)	No. of bets	Returns (£)	Profit/ Loss (£)	No. of bets	Returns (£)	Profit/Loss (£)
1%	358	356.24	-0.49%	280	266.25	-4.91%	284	276.04	-2.80%
2%	325	320.21	-1.47%	240	225.93	-5.86%	234	235.98	0.85%
3%	275	277.85	1.04%	189	187.07	-1.02%	192	191.12	-0.46%
4%	225	236.87	5.28%	136	144.85	6.51%	147	159.44	8.46%
5%	169	183.19	8.40%	109	112.13	2.87%	123	134.66	9.48%
6%	131	148.4	13.28%	85	84.96	-0.05%	95	102.31	7.69%
7%	107	119.92	12.07%	68	64.86	-4.62%	67	68.91	2.85%
8%	84	92.43	10.04%	53	54.79	3.38%	45	49.53	10.07%
9%	71	82.36	16.00%	36	39.19	8.86%	34	32.71	-3.79%
10%	52	62.61	20.40%	26	16.97	-34.73%	24	23.55	-1.88%
11%	41	55.61	35.63%	15	7.82	-47.87%	19	21.82	14.84%
12%	25	18.05	-27.80%	12	7.82	-34.83%	13	7.82	-39.85%
13%	15	10.39	-30.73%	10	7.82	-21.80%	10	7.82	-21.80%
14%	12	8.3	-30.83%	8	7.82	-2.25%	10	7.82	-21.80%
15%	10	8.3	-17.00%	7	7.82	11.71%	7	7.82	11.71%
16%	7	8.3	18.57%	5	6.2	24.00%	6	6.2	3.33%
17%	6	8.3	38.33%	2	0	-100%	3	2.4	-20.00%
18%	5	5.9	18.00%	2	0	-100%	2	0	-100%
19%	2	0	-100%	1	0	-100%	1	0	-100%
20%	2	0	-100%	1	0	-100%	1	0	-100%

Appendix B.7: Forecast examples generated by pi-football

Table B.7.1. Objective (f_O) and subjective (f_S) forecasts generated by pi-football, at the beginning of the EPL season 2010/11

	Home	Away		Objective (f_O)			Subjective (f_S)		
Date	Team	Team	Result	$p(H)$	$p(D)$	$p(A)$	$p(H)$	$p(D)$	$p(A)$
14/08/2010	Aston Villa	West Ham	H	60.92	23.971	15.109	61.735	23.67	14.596
14/08/2010	Blackburn	Everton	H	34.382	29.314	36.304	36.338	29.781	33.881
14/08/2010	Bolton	Fulham	D	46.863	29.199	23.938	46.863	29.199	23.938
14/08/2010	Chelsea	West Brom	H	87.055	12.706	0.24	89.581	10.227	0.192
14/08/2010	Sunderland	Birmingham	D	44.366	29.623	26.011	44.197	29.679	26.124
14/08/2010	Tottenham	Man City	D	35.178	33.654	31.168	32.82	33.756	33.424
14/08/2010	Wigan	Blackpool	A	53.939	30.156	15.905	53.939	30.156	15.905
14/08/2010	Wolves	Stoke	H	38.763	31.563	29.674	37.778	31.746	30.477
15/08/2010	Liverpool	Arsenal	D	51.705	27.305	20.99	54.007	26.773	19.22
16/08/2010	Man United	Newcastle	H	81.665	16.058	2.277	83.853	14.18	1.966
21/08/2010	Arsenal	Blackpool	H	85.569	12.668	1.763	85.695	12.56	1.746
21/08/2010	Birmingham	Blackburn	H	44.269	29.088	26.643	49.695	28.632	21.673
21/08/2010	Everton	Wolves	D	73.202	17.433	9.365	69.731	20.077	10.192
21/08/2010	Stoke	Tottenham	A	27.657	29.283	43.059	28.289	29.58	42.13
21/08/2010	West Brom	Sunderland	H	36.848	33.163	29.989	36.325	33.216	30.459
21/08/2010	West Ham	Bolton	A	39.606	32.217	28.177	35.012	33.074	31.913
21/08/2010	Wigan	Chelsea	A	9.945	16.713	73.342	6.465	14.345	79.19
22/08/2010	Fulham	Man United	D	13.416	22.345	64.239	12.059	21.442	66.499
22/08/2010	Newcastle	Aston Villa	H	26.934	31.612	41.455	38.277	35.144	26.58
23/08/2010	Man City	Liverpool	H	55.566	26.104	18.33	59.331	24.983	15.686
28/08/2010	Blackburn	Arsenal	A	29.444	31.547	39.009	24.496	31.194	44.31
28/08/2010	Blackpool	Fulham	D	28.052	31.672	40.276	28.272	31.732	39.996
28/08/2010	Chelsea	Stoke	H	80.673	16.736	2.591	84.022	13.905	2.073
28/08/2010	Man United	West Ham	H	82.525	15.553	1.922	84.627	13.711	1.662
28/08/2010	Tottenham	Wigan	A	73.716	17.443	8.841	73.327	17.74	8.934
28/08/2010	Wolves	Newcastle	D	40.609	32.837	26.554	37.192	33.491	29.318
29/08/2010	Aston Villa	Everton	H	45.276	31.446	23.277	44.676	31.63	23.695
29/08/2010	Bolton	Birmingham	D	39.858	31.208	28.934	36.146	32.013	31.84
29/08/2010	Liverpool	West Brom	H	80.318	15.187	4.495	77.822	17.212	4.967
29/08/2010	Sunderland	Man City	H	21.155	20.44	58.405	21.584	21.237	57.179
11/09/2010	Arsenal	Bolton	H	70.745	19.864	9.391	70.751	19.861	9.388
11/09/2010	Everton	Man United	D	27.891	25.825	46.284	31.386	28.593	40.021
11/09/2010	Fulham	Wolves	H	46.98	29.379	23.641	48.281	29.125	22.594
11/09/2010	Man City	Blackburn	D	69.118	20.636	10.246	62.251	25.453	12.296
11/09/2010	Newcastle	Blackpool	A	55.782	31.301	12.918	51.035	33.384	15.581
11/09/2010	West Brom	Tottenham	D	22.674	28.013	49.314	25.911	30.475	43.614
11/09/2010	West Ham	Chelsea	A	7.98	16.013	76.007	7.879	15.911	76.21
11/09/2010	Wigan	Sunderland	D	40.77	32.102	27.128	41.178	32.039	26.784
12/09/2010	Birmingham	Liverpool	D	30.374	29.364	40.262	35.557	31.287	33.155
13/09/2010	Stoke	Aston Villa	H	29.946	29.846	40.208	35.597	31.808	32.595
18/09/2010	Aston Villa	Bolton	D	67.813	20.418	11.768	66.943	21.027	12.03
18/09/2010	Blackburn	Fulham	D	49.733	28.365	21.902	48.58	28.861	22.559
18/09/2010	Everton	Newcastle	A	64.358	22.042	13.6	63.488	22.615	13.898
18/09/2010	Stoke	West Ham	D	45.372	31.286	23.342	39.697	33.048	27.255
18/09/2010	Sunderland	Arsenal	D	17.051	20.505	62.444	21.997	30.62	47.383
18/09/2010	Tottenham	Wolves	H	74.223	17.506	8.271	43.964	40.629	15.407
18/09/2010	West Brom	Birmingham	H	33.397	32.167	34.436	34.729	32.261	33.01
19/09/2010	Chelsea	Blackpool	H	88.112	11.363	0.525	88.753	10.751	0.496
19/09/2010	Man United	Liverpool	H	58.15	28.169	13.681	61.165	26.618	12.217
19/09/2010	Wigan	Man City	A	23.721	26.167	50.113	25.023	27.358	47.619

APPENDIX C

Profiting from an inefficient Association Football gambling market: Prediction, risk and uncertainty using Bayesian networks (Chapter 7)

Appendix C.1: Cumulative Returns based on BP_1 and BP_2

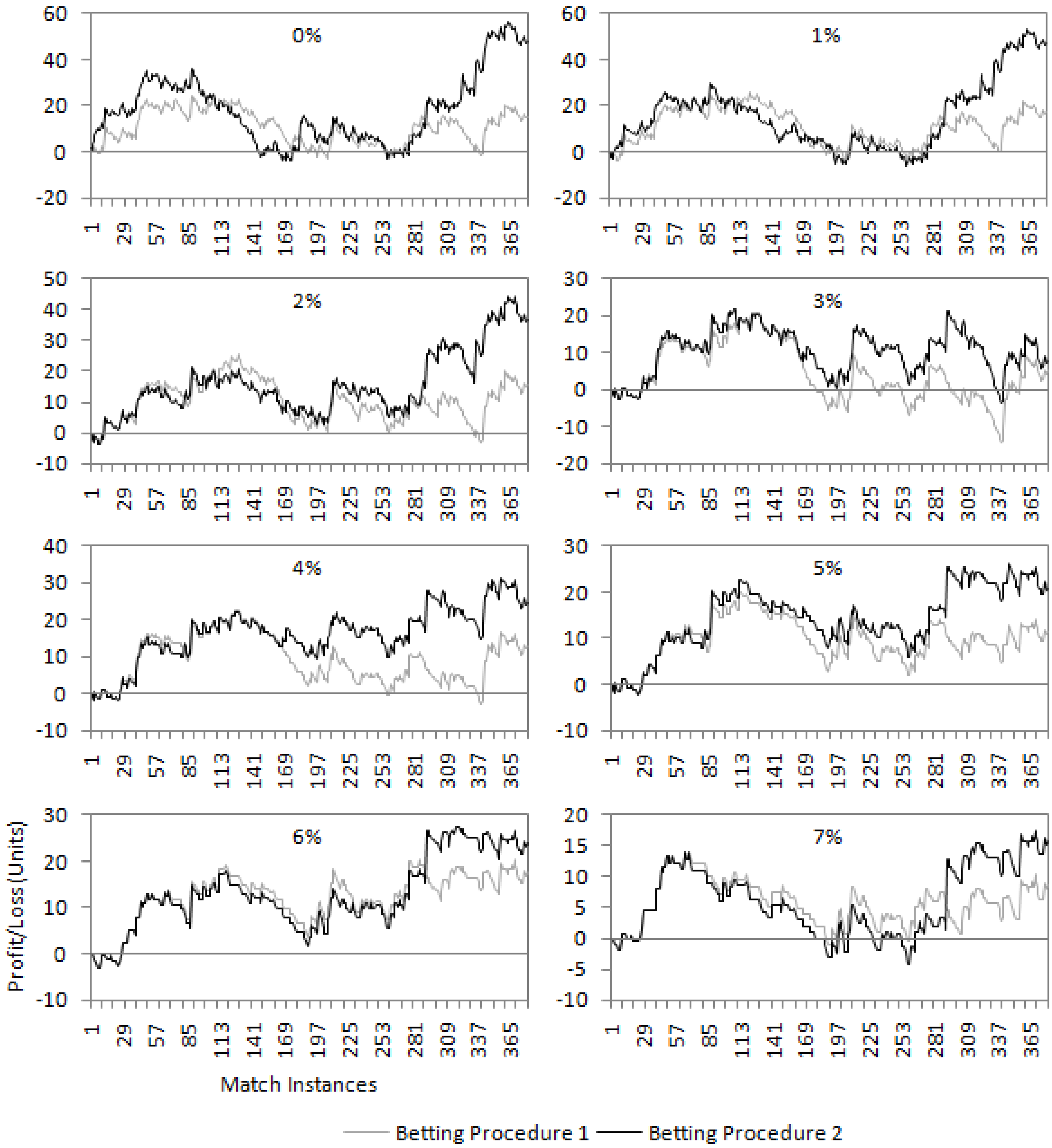


Figure C.1.1. Cumulative unit-based returns based on BP_1 and BP_2 according to the specified discrepancy level.

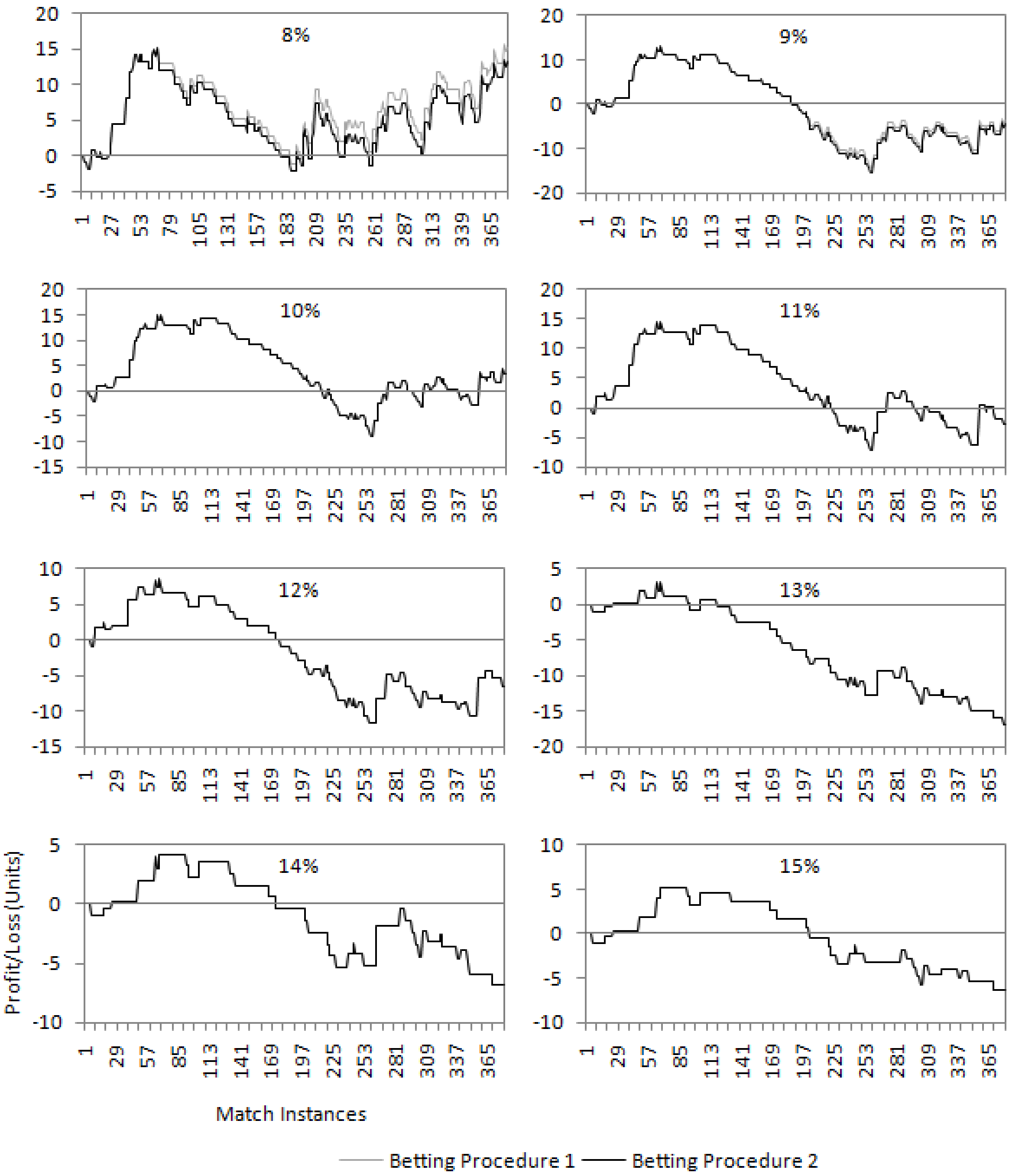


Figure C.1.2. Cumulative unit-based returns based on BP_1 and BP_2 according to the specified discrepancy level.

Appendix C.2: Risk Assessment of Profit and Loss based on the specified betting procedure.

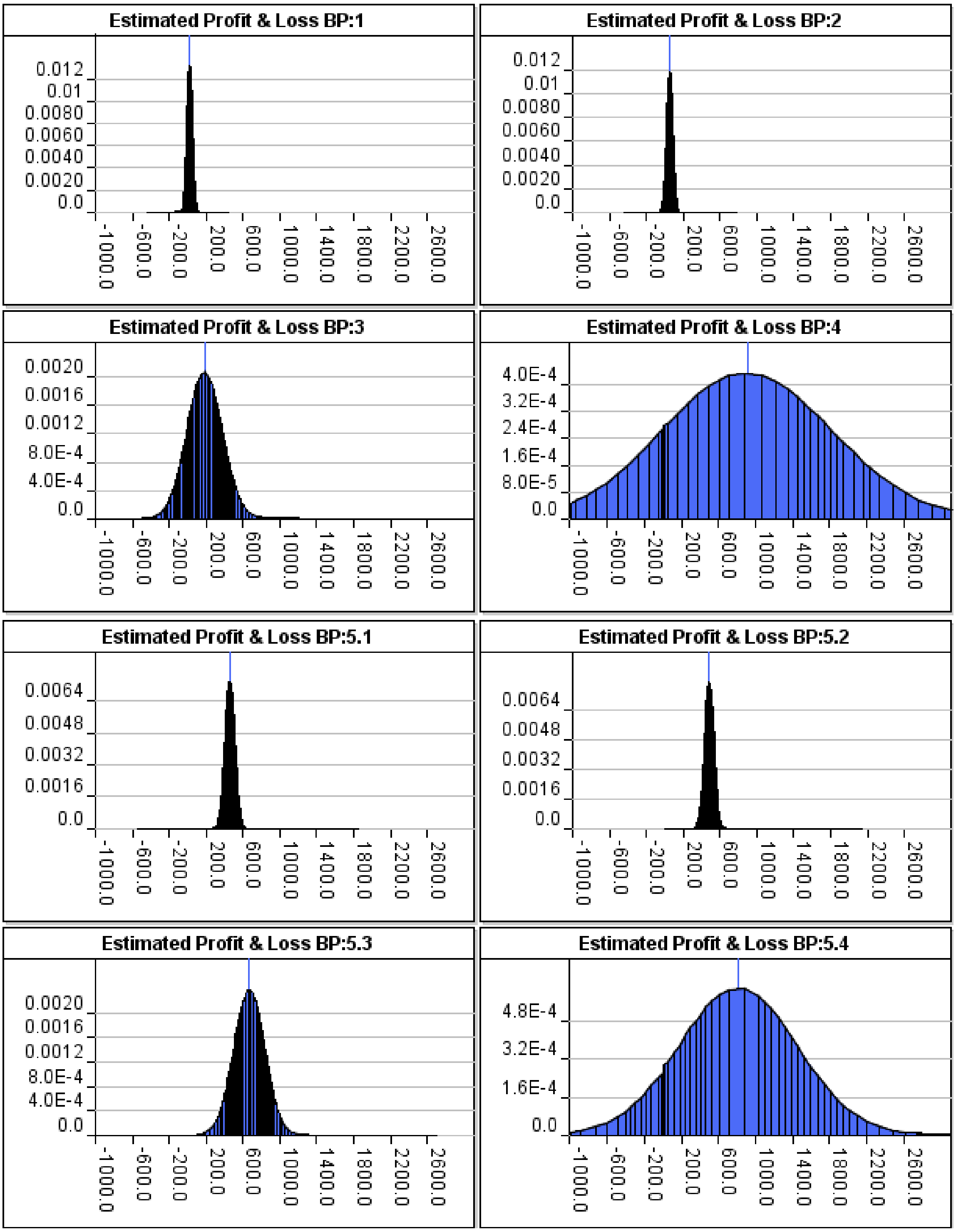


Figure C2.1. Risk assessment of concluding expected season returns according to each betting procedure.

Appendix C.3: Model performance when considering arbitrage opportunities.

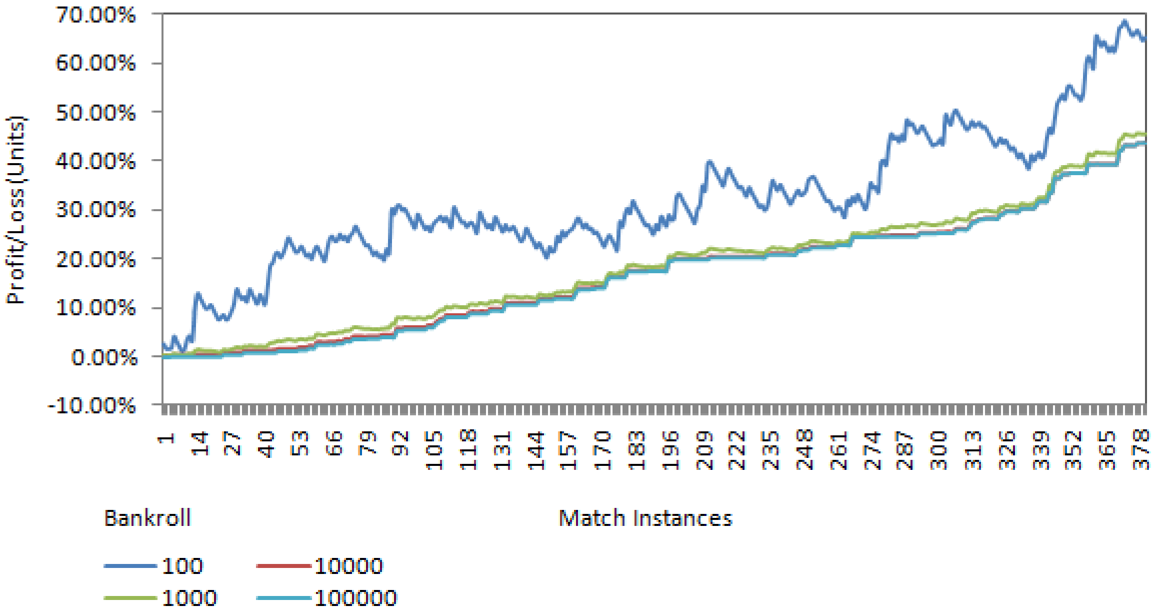


Figure C.3.1. Cumulative unit-based returns based on $BP_{5.1}$ assuming no discrepancy restrictions (set to 0%) and according to the specified bankrolls prior to initialising the betting simulation.

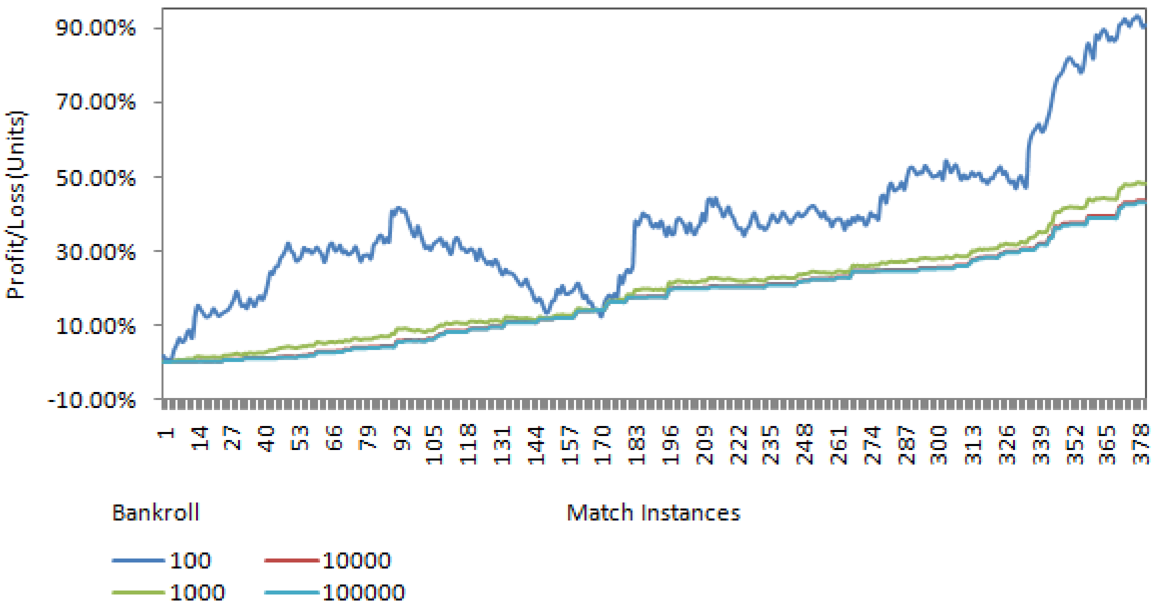


Figure C.3.2. Cumulative unit-based returns based on $BP_{5.2}$ assuming no discrepancy restrictions (set to 0%) and according to the specified bankrolls prior to initialising the betting simulation.

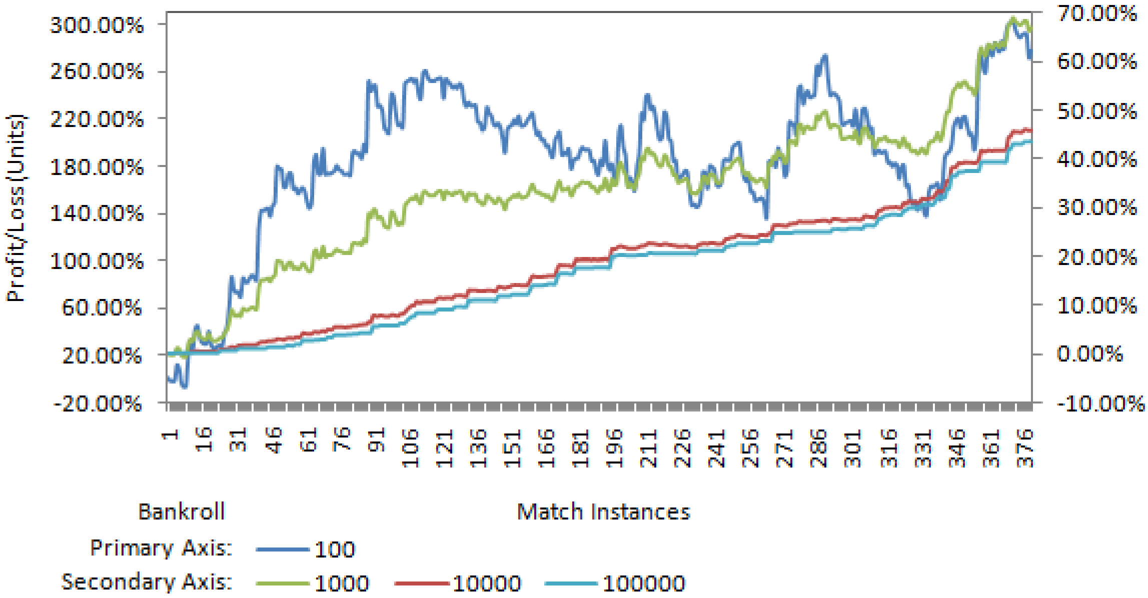


Figure C.3.3. Cumulative unit-based returns based on $BP_{5.3}$ and according to the specified bankrolls prior to initialising the betting simulation.

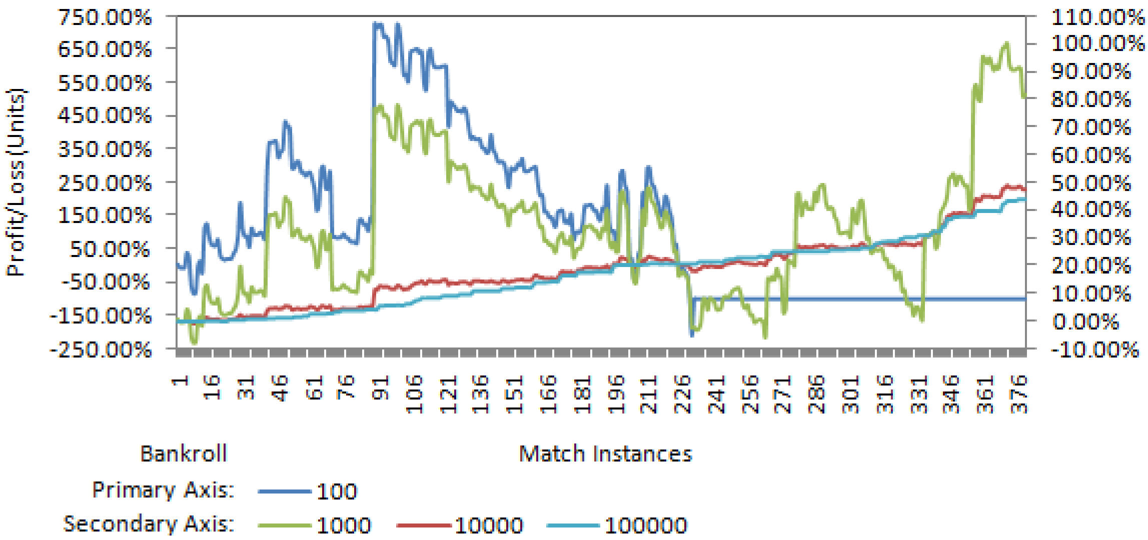


Figure C.3.4. Cumulative unit-based returns based on $BP_{5.4}$ and according to the specified bankrolls prior to initialising the betting simulation.

Appendix C.4: Performance based on parameters of component level 3

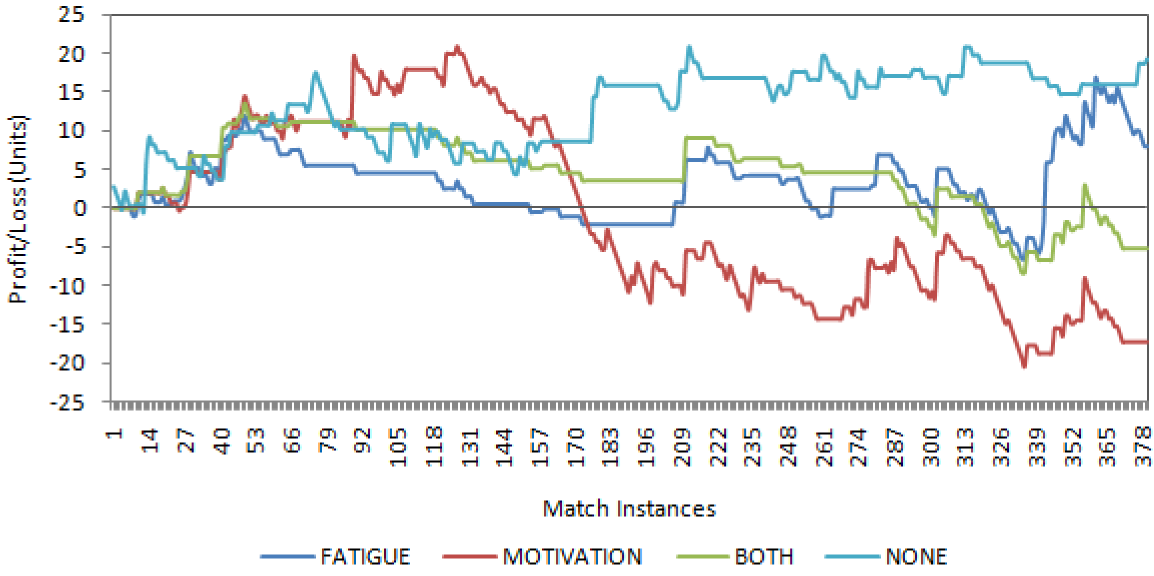


Figure C.4.1. Cumulative unit-based returns based on BP_1 for match instances with the specified evidence.

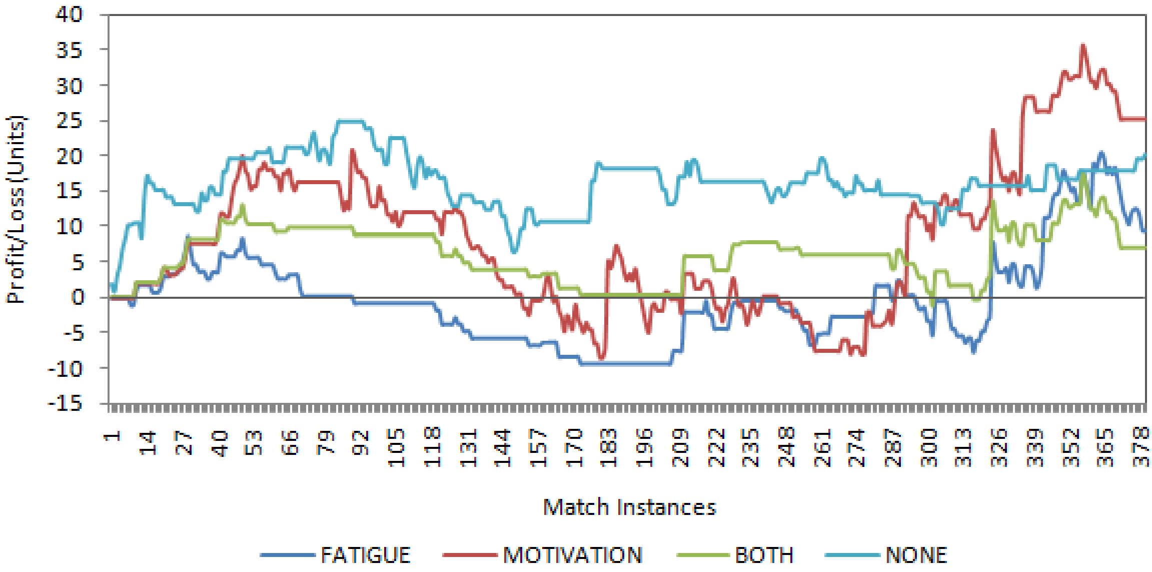


Figure C.4.2. Cumulative unit-based returns based on BP_2 for match instances with the specified evidence.

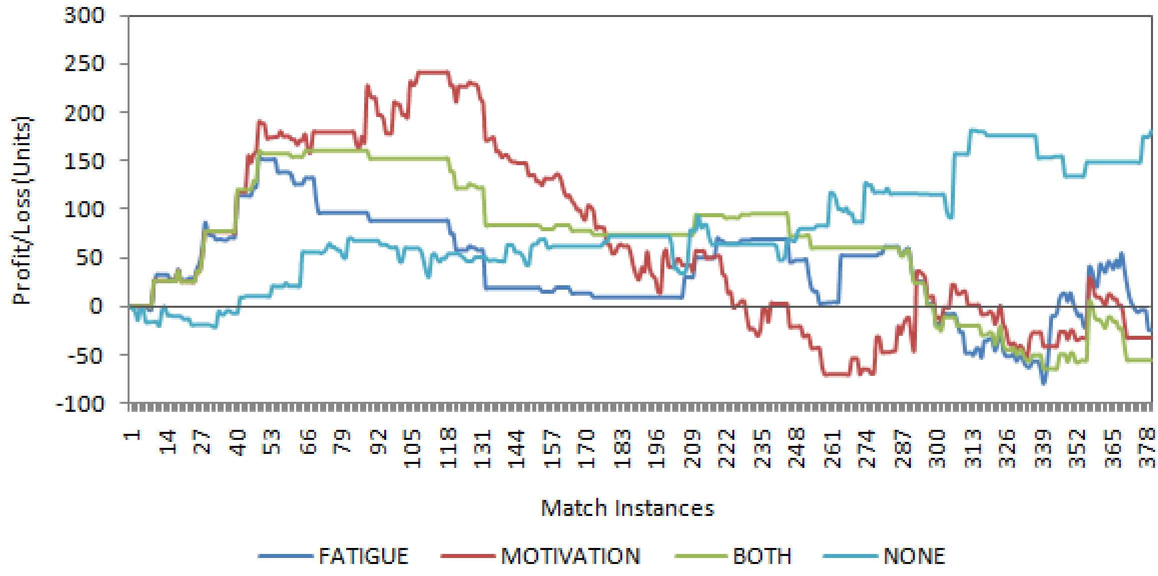


Figure C.4.3. Cumulative unit-based returns based on BP_3 for match instances with the specified evidence.

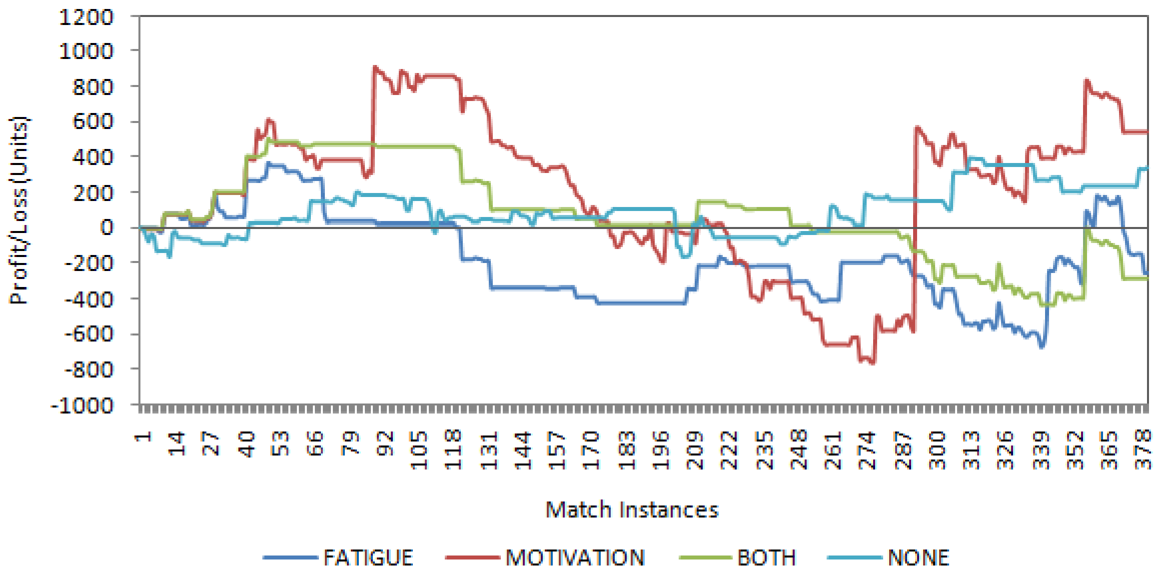


Figure C.4.4. Cumulative unit-based returns based on BP_4 for match instances with the specified evidence.

Appendix C.5: Team-based efficiency against market odds.

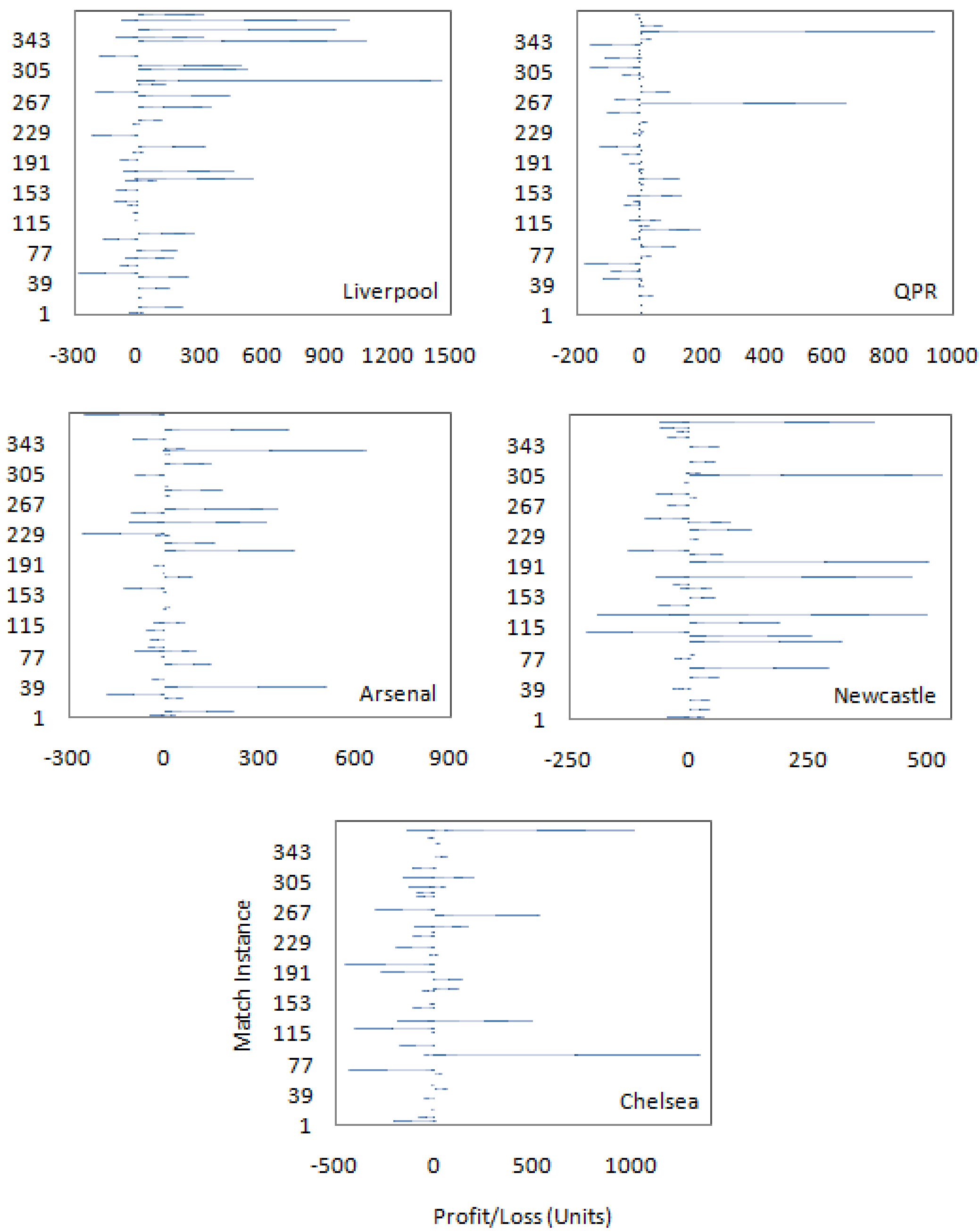


Figure C.5.1. Team-based explicit returns against market odds throughout the EPL season.

Appendix C.6: Unit-based performance relative to the previous model

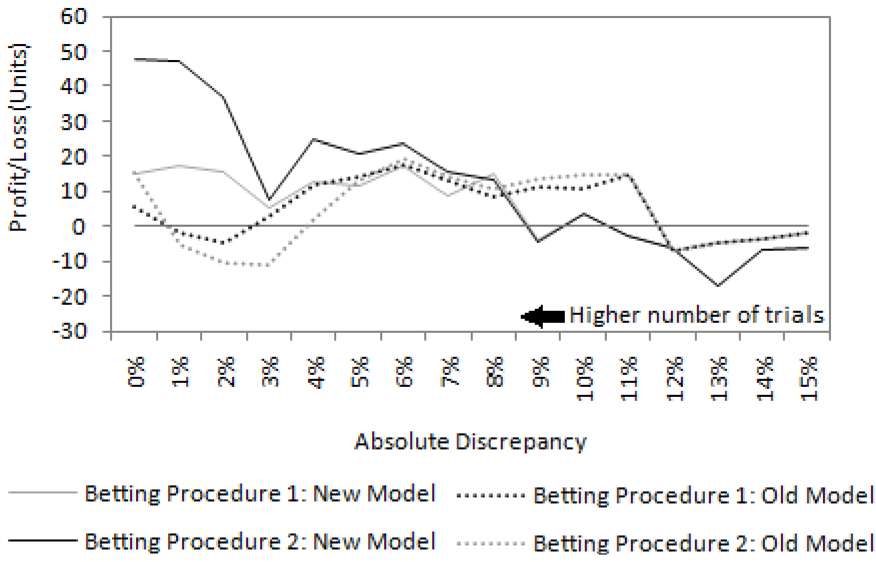


Figure C.6.1. Cumulative unit-based returns based on BP_1 and BP_2; a comparison between the new and the old model.

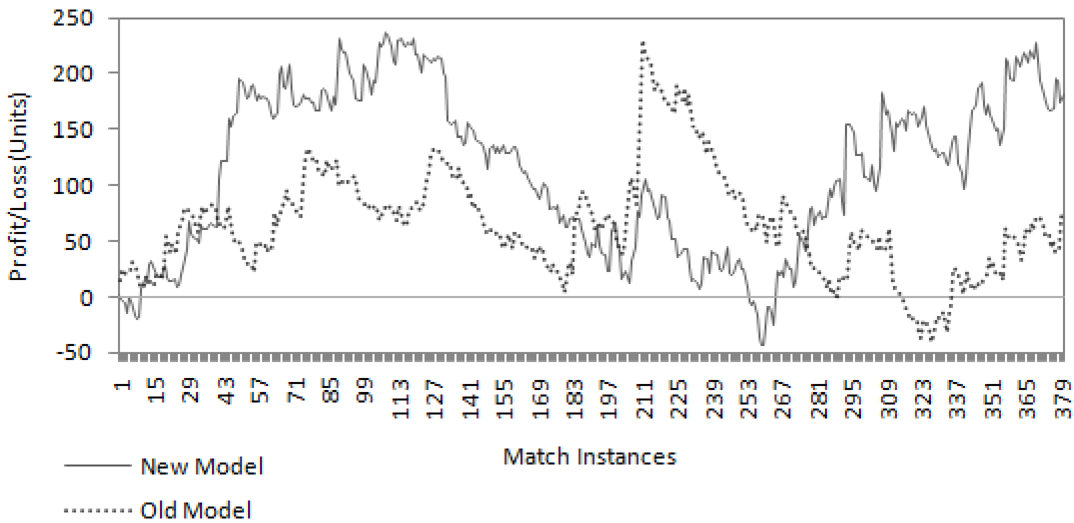


Figure C.6.2. Cumulative unit-based returns based on BP_3; a comparison between the new and the old model.

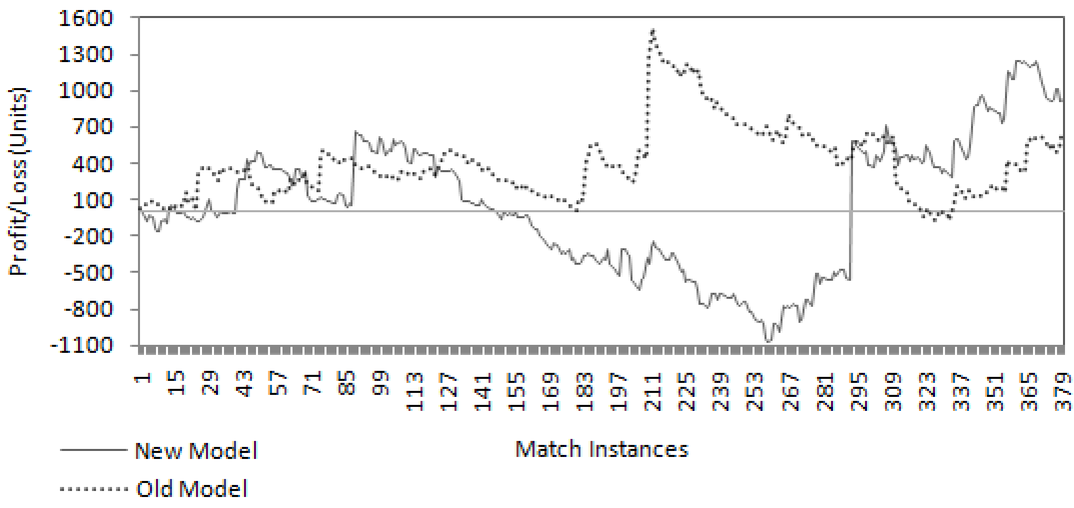


Figure C.6.3. Cumulative unit-based returns based on BP_4; a comparison between the new and the old model.

APPENDIX D

Determining the level of ability of football teams by dynamic ratings based on the relative discrepancies in scores between adversaries (Chapter 8)

Appendix D.1: Rating development over a period of 20 seasons

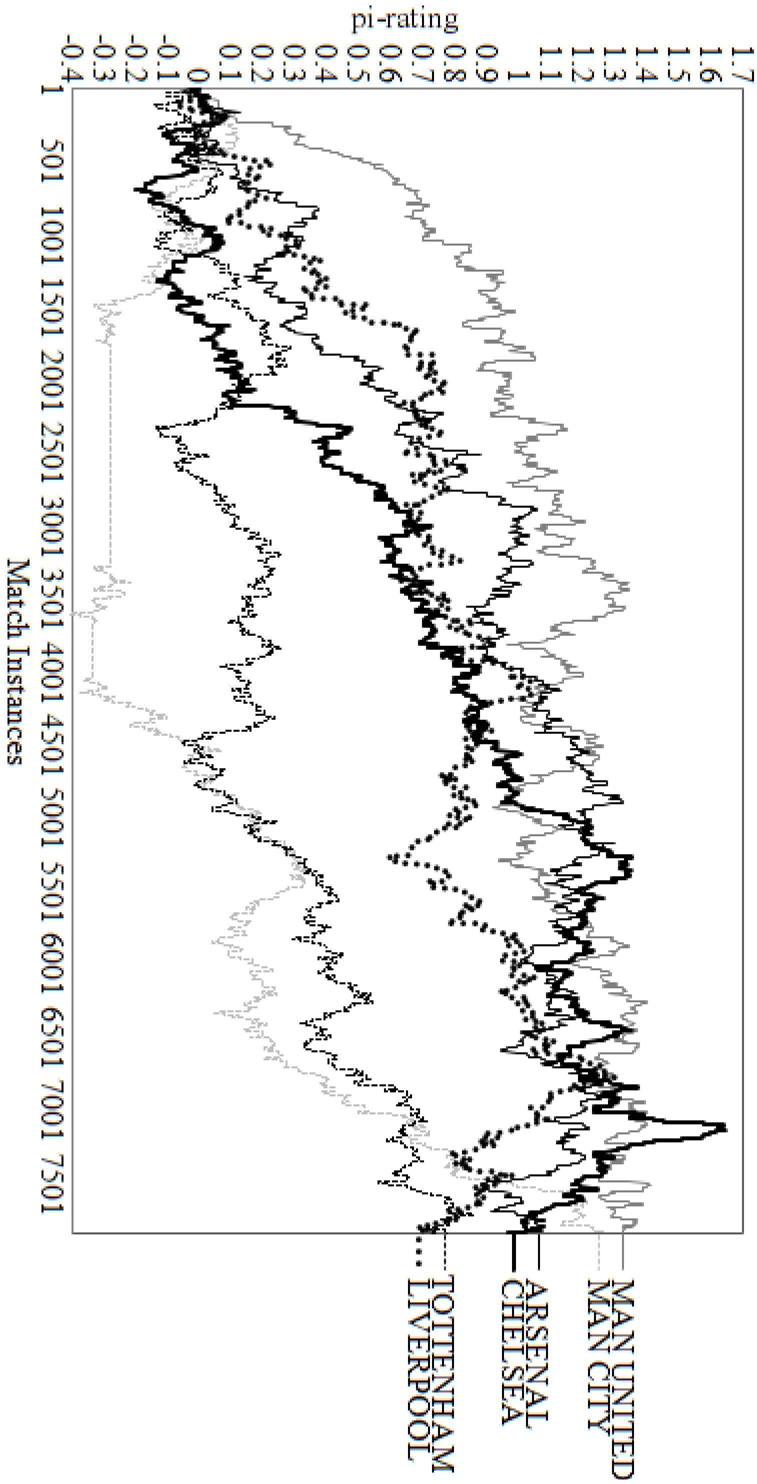


Figure D.1.1. Rating development over a period of 20 seasons for the six most popular EPL teams (from season 1992/93 to season 2011/12 inclusive).

References

[1] Agena. (2012). Agena: Bayesian Network and Simulation Software for Risk Analysis and Decision Support. Retrieved January 26, 2012, from http://www.agenarisk.com

[2] Agena. (2009). AgenaRisk 5.0 User Manual.Agena Ltd, pp. 195.

[3] Ali, M. M. (1979). Some evidence of the efficiency of a speculative market. Econometrica, 47, 387-392.

[4] Ali, M. (1977). Probability and utility estimates for racetrack bettors. Journal of Political Economy, 85, 803-815.

[5] Armstrong, J. (2007a). Significance tests harm progress in forecasting. International Journal of Forecasting , 23: 321-327.

[6] Armstrong, J. (2007b). Statistical Significance Tests are Unnecessary Even When Properly Done. International Journal of Forecasting , 23: 335-336.

[7] Armstrong, J., and Collopy, F. (1992). Error Measures For Generalizing About Forecasting Methods: Empirical Comparisons. International Journal of Forecasting, 8, 69-80.

[8] AT Sigma (2007). Advanced Technologies: AT Sigma SPC Master Suite. Retrieved June 29, 2012, from http://www.atsigma.com/

[9] Asch, P., Malkiel, B. G., and Quandt, R. E. (1984). Market efficiency in racetrack betting. Journal of Business, 57, 165-175.

[10] Baio, G., and Blangiardo, M. (2010). Bayesian hierarchical model for the prediction of football results. Journal of Applied Statistics, 37:2, 253-264.

[11] Bangsø, O., Flores, M. J., Jensen, F. V. (2003). Plug & Play Obejct Oriented Bayesian Networks. Proceedings of the Tenth Conference of the Spanish Association for Artificial Intelligence. Spain.

[12] Bayes, T. (1763). "An Essay Towards Solving a Problem in the Doctrine of Chances." Philosophical Transaction of the Royal Society of London, 53: 370-418.

[13] Bayes Server. (2012) What is Bayes Server? Bayes Server - Bayesian network software. Retrieved July 20, 2012, from http://www.bayesserver.com/

[14] Bayesia. (2001) Bayesia - Make the Best Decisions. Retrieved July 20, 2012, from http://www.bayesia.com/

[15] Bayesian Knowledge Discoverer. (1998) Bayesian Knowledge Discoverer. Retrieved July 20, 2012, from http://projects.kmi.open.ac.uk/bkd/pages/bkd.html

[16] Berger, J.O., and Sellke, T. (1987). Testing a point null hypothesis: The irreconcilability of p values and evidence. Journal of the American Statistical Association, 82, 112–139.

[17] betfair. (2000). About Us: How does Betfair work? Retrieved November 22, 2011, from http://help.betfair.com/contents/itemId/i65767197/index.en.html

[18] BNET. (2004) bnet.builder. BNET, Belief Network Tools from Charles River Analytics. Retrieved July 20, 2012, from https://www.cra.com/commercial-solutions/bnet-builder.asp

[19] Brier, G. (1950). Verification of forecasts expressed in terms of probability. Monthly Weather Review, 78, 1-3.

[20] Buchdahl, J. (2003). Fixed Odds Sports Betting: Statistical Forecasting and Risk Management. High Stakes.

[21] Buchner, A., Dubitzky, W., Schuster, A., Lopes, P., O'Doneghue, P., Hughes, J., et al. (1997). Corporate evidential decision making in performance prediction domains. Proceedings of the Thirteenth Conference on Uncertainty in Artificial Intelligence (UAI'97). Brown University, Providence, Rhode Island, USA.

[22] bwin Group. (2009). Annual Report 2008. Vienna, Austria: bwin Interactive Entertainment AG.

[23] bwin Group. (2010). Annual Financial Report 2009. Vienna, Austria: bwin Interactive Entertainment AG.

[24] Cain, M., Law, D., and Peel, D. (2000). The Favourite-Loghshot Bias and Market Efficiency in UK Football Betting. Scottish Journal of Political Economy, 47, 25-36.

[25] Casscells, W., Schoenberger, A., and Grayboys, T. (1978). Interpretation by physicians of clinical laboratory results. New England Journal of Medicine, 299, 999-1000.

[26] Castagna, C., Impellizzeri, F. M., Chamari, K., Carlomagno, D., and Rampinini E. (2006). Aerobic fitness and yo-yo continuous and intermittent tests performances in soccer players: a correlation study. Journal of Strength & Conditioning Research, 20:320-325.

[27] Cheng, Jie. (2001). J Cheng's Bayesian Belief Network Software. Retrieved July 20, 2012, from http://webdocs.cs.ualberta.ca/~jcheng/bnsoft.htm

[28] Cheng, J., Bell, D. A. and Liu, W. (1997). An algorithm for Bayesian belief network construction from data. In Proceedings of the Sixth International Workshop on Artificial Intelligence and Statistics, Smyth, P. & Madigan, D. (eds). Fort Lauderdale, USA, 83–90.

[29] Clarke, S. R., and Norman, J. M. (1995). Home ground advantage of individual clubs in English soccer. The Statistician, 44, 509–521.

[30] Cohen, J. (1990). Things I Have Learned (So Far). American Psychological Association.

[31] Constantinou, Anthony C. and Fenton, Norman E. (2012a). 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, DOI: 10.1515/1559-0410.1418 (Draft available online at: http://constantinou.info/downloads/papers/solvingtheproblem.pdf)

[32] Constantinou, Anthony C. and Fenton, Norman E. (2012b). Evidence of an (intended) inefficient Association Football gambling market. Submitted for publication. (Draft available online at: http://constantinou.info/downloads/papers/evidenceofinefficiency.pdf)

[33] Constantinou, Anthony C. and Fenton, Norman E. (2012c). Determining the level of ability of football teams by dynamic ratings based on the relative discrepancies in sores between adversaries. Submitted for publication. (Draft available online at: http://www.constantinou.info/downloads/papers/pi-ratings.pdf)

[34] Constantinou, Anthony C., Fenton, Norman E. and Neil, Martin. (2012a). pi-football: A Bayesian network model for forecasting Association Football match outcomes. To

appear in Knowledge-Based Systems, 2012. Online publication August 21, 2012, DOI: 10.1016/j.knosys.2012.07.008 (Draft available online at: http://constantinou.info/downloads/papers/pi-model11.pdf)

[35] Constantinou, Anthony C., Fenton, Norman E. and Neil, Martin. (2012b). Profiting for an inefficient Association Football gambling market: Prediction, Risk and Uncertainty using Bayesian networks. Submitted for publication, 2012. (Draft available online at: http://constantinou.info/downloads/papers/pi-model12.pdf)

[36] Corani, G., Zaffalon, M. (2008). Learning Reliable Classifiers From Small or Incomplete Data Sets: The Naive Credal Classifier 2. Journal of Machine Learning Research, 9, 581--621.

[37] Crowder, M., Dixon, M., Ledford, A., and Robinson, M. (2002). Dynamic modelling and prediction of English football league matches for betting. The Statistician, 51, 157– 168.

[38] Cumming, G. (2005). Understanding the average probability of replication: Comment on Killeen (2005). Psychological Science, 16, 1002–1004.

[39] Daly, R., Shen, Q., and Aitken, S. (2011) Learning Bayesian networks: approaches and issues. The Knowledge Engineering Review, 26 (2). Pp 99-157.

[40] Darwiche, A. 2002. A logical approach to factoring belief networks. In Proceedings of the Eight International Conference on Principles of Knowledge Representation and Reasoning (KR-02), Fensel, D., Giunchiglia, F., McGuinness, D. L. & Williams, M.-A. (eds). Morgan Kaufmann, 409–420.

[41] Davis, G. A. (2003). Bayesian reconstruction of traffic accidents. Law, Probability and Risk 2 (2): 69–89.

[42] de Campos, L. M., Fernández-Luna, J. M., and Huete, J. F. (2004). Bayesian networks and information retrieval: an introduction to the special issue. Information Processing & Management (Elsevier) 40 (5): 727–733.

[43] Decision Systems Laboratory. (2005). GeNle & SMILE. About GeNle & SMILE. Retrieved July 20, 2012, from http://genie.sis.pitt.edu/about.html

[44] Díez,, F. J., Mira, J., Iturralde, E., and Zubillaga, S. (1997). DIAVAL, a Bayesian expert system for echocardiography. Artificial Intelligence in Medicine (Elsevier) 10 (1): 59–73.

[45] Dixon, M., and Coles, S. (1997). Modelling association football scores and inefficiencies in the football betting market. Applied Statistics, 46, 265-80.

[46] Dixon, M., and Pope, P. (2004). The value of statistical forecasts in the UK association football betting market. International Journal of Forecasting, 20, 697-711.

[47] Doros, G., and Geier, A.B. (2005). Probability of replication revisited: Comment on ‘‘An alternative to null-hypothesis significance tests.’’ Psychological Science, 16, 1005–1006.

[48] Douglas, G. A., and Bland, J. M. (1995). Statistics notes: "Absence of evidence is not evidence of absence". BMJ , 311:485.

[49] Dunning, E. G., and Joseph A. Maguire, R. E. (1993). The Sports Process: A Comparative and Developmental Approach. Champaign: Human Kinetics, p. 129.

[50] Dunning, E. (1999). Sport Matters: Sociological Studies of Sport, Violence and Civilisation. London: Routledge.

[51] Edwards, W., Lindman, H., and Savage, L.J. (1963). Bayesian statistical inference for psychological research. Psychological Review, 70, 193–242.

[52] Efron, B. (2005). "Bayesians, Frequentists, and Scientists." Journal of the American Statistical Association, 100: 1-5.

[53] Elo, A. E. (1978). The rating of chess players, past and present. New York: Arco Publishing.

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

[55] Falter, M., and Perignnon, C. (2000). Demand for football and intra-match winning probability: an essay on the glorious uncertainty of sports. Applied Economics , 32, 1757-1765.

[56] FDEP. (2001). A Program for Inducing Functional Dependencies from Relations. FDEP Home Page. Retrieved July 20, 2012, from http://www.cs.bris.ac.uk/~flach/fdep/

[57] Fenton, N. E., Neil, M. and Caballero, J. G. (2007). Using Ranked nodes to model qualitative judgments in Bayesian Networks. IEEE TKDE 19(10), 1420-1432

[58] Fenton, N.E. and M. Neil. (2012). Risk Assessment and Decision Analysis with Bayesian Networks. CRC Press.

[59] FIFA. (2012). FIFA. Retrieved March 27, 2012, from FIFA/Coca-Cola World Ranking Procedure:

http://www.fifa.com/worldranking/procedureandschedule/menprocedure/index.html

[60] Fildes, R., and Stekler, H. (2002). The state of macroeconomic forecasting. Journal of Macroeconomics, 24, 435-468.

[61] Finnigan, M., and Nordsted, P. (2010). The Premier Football Betting Handbook 2010/11. Hampshire, Great Britain: Harriman House.

[62] Fisher, R. (1922). On the interpretation of χ^2 from contingency tables, and the calculation of P. Journal of Royal Statistical Society , 85 (1): 87-94.

[63] Fisher, R. (1954). Statistical Methods for Research Workers. Oliver and Boyd.

[64] Flach, Peter A. and Savnik, Iztok. (1999). Database dependency discovery: a machine learning approach. AI Communications, 12(3):139-160.

[65] Football-Data. (n.d.). Football-Data.co.uk. Retrieved August 25, 2011, from Football Results, Statistics & Soccer Betting Odds Data: http://www.football-data.co.uk/englandm.php

[66] Forrest, D., and Simmons, R. (2000). Forecasting sport: the behaviour and performance of football tipsters. International Journal of Forecasting, 16, 317-331.

[67] Forrest, D., and Simmons, R. (2001). Globalisation and efficiency in the fixed-odds soccer betting market. University of Salford, Salford: Mimeo.

[68] Forrest, D., and Simmons, R. (2002). Outcome uncertainty and attendance demand in sport: the case of English soccer. The Statistician, 2, 241-291.

[69] Forrest, D., and Simmons, R. (2008). Sentiment in the betting market on Spanish football. Applied Economics, 40, 119-126.

[70] Forrest, D., Goddard, J., and Simmons, R. (2005). Odds-setters as forecasters: The case of English football. International Journal of Forecasting, 21, 551-564.

[71] Friedman, N.F., Linial, M., Nachman, I., and Pe'er, D. (2000). Using Bayesian Networks to Analyze Expression Data. Journal of Computational Biology (Larchmont, New York: Mary Ann Liebert, Inc.) 7 (3/4): 601–620

[72] Garthwaite, P., Kadane, J., and O'Hagan, A. (2005). Statistical Methods for Eliciting Probability Distributions. Journal of the American Statistical Association, 100, 680-700.

[73] Gigerenzer, G. (2002). Reckoning with Risk: Learning to Live with Uncertainty, Penguin Books.

[74] Global Betting and Gaming Consultants. (2001). 1st annual review of the global betting and gaming market, 2001. West Bromwich: Global Betting and Gaming Consultants.

[75] Gneiting, T., and Raftery, A. (2007). Strictly Proper Scoring Rules, Prediction, and Estimation. Journal of the American Statistical Association, 102(477), 359-378.

[76] Goddard, J. (2005). Regression models for forecasting goals and match results in association football. International Journal of Forecasting, 21, 331-340.

[77] Goddard, J., and Asimakopoulos, I. (2004). Forecasting Football Results and the Efficiency of Fixed-odds Betting. Journal of Forecasting, 23, 51-66.

[78] Goldberg, A. and Robson, D. (1983). Smalltalk-80: The Language and its Implementation. Addison-Wesley, 1983.

[79] Golec, J., and Tamarkin, M. (1991). The degree of inefficiency in the football betting market. Journal of Financial Economics, 30, 311-323.

[80] Golec, J., and Tamarkin, M. (1998). Bettors love skewness, not risk, at the horse track. Journal of Political Economy, 106, 205-225.

[81] Graham, I., and Stott, H. (2008). Predicting bookmaker odds and efficiency for UK football. Applied Ecomonics, 40, 99-109.

[82] Haigh, J. (2003). Taking Chances: Winning with Probability, Oxford University Press.

[83] Halicioglu, F. (2005a). Can we predict the outcome of the international football tournaments? : the case of Euro 2000. Doğuş Üniversitesi Dergisi , 6, 112-122.

[84] Halicioglu, F. (2005b). Forecasting the Professional Team Sporting Events: Evidence from Euro 2000 and 2004 Football Tournaments. 5th International Conference on Sports and Culture: Economic, Management and Marketing Aspects, (pp. 30-31). Athens, Greece.

[85] Harville, D. A. (1977) The use of linear-model methodology to rate high school or college football teams. Journal of the American Statistical Association. 72, 278-289.25583854

[86] Hausch, D. B., Ziemba, W. T., and Rubinstein, M. (1981). Efficiency of the market for racetrack betting. Management Science, 27, 1435-1452.

[87] Heckerman D. (1995a). A Bayesian Approach to Learning Causal Networks. Microsoft Research technical report, MSR-TR-95-04. Morgan Kaufmann.

[88] Heckerman D. (1995b). A Tutorial on Learning With Bayesian Networks. Microsoft Technical Report, MSR-TR-95-06. Microsoft Research.

[89] Heckerman D., Mamdani, A., Wellman, M. (1995). Real-world applications of Bayesian networks. Comm ACM, vol. 38, no. 3, pp. 25-26.

[90] Heckerman, D. 2007. A Bayesian approach to learning causal networks. In Advances in Decision Analysis: from Foundations to Applications, Edwards, W. & Miles R. F. Jr (eds). Chapter 11, Cambridge University Press, 202–220.

[91] Hendry, D. (1997). The Econometrics of Macroeconomic Forecasting. The Economic Journal, 1330-1357.

[92] Henery, R. J. (1999). Measures of over-round in performance index betting. The Statistician, 48, 435-439.

[93] Hirotsu, N., and Wright, M. (2003). An evaluation of characteristics of teams in association football by using a Markov process model. The Statistician, 52: 4, 591-602.

[94] Hsu, William H. (2004). Bayesian Network tools in Java. Bayesian Network tools in Java (BNJ) - Kansas State University Lab for Knowledge Discovery in Databases. Retrieved July 20, 2012, from http://bnj.sourceforge.net/

[95] HUGIN EXPERT. (1989). HUGIN EXPERT, The leading decision support tool. HUGIN EXPERT - Advanced Decision Support using Bayesian Networks and Influence Diagrams. Retrieved July 20, 2012, from http://www.hugin.com/

[96] Hvattum, L. M., and Arntzen, H. (2010). Using ELO ratings for match result prediction in association football. International Journal of Forecasting, 26, 460-470.

[97] Ioannidis, J. P. (2005a). Contradicted and initially stronger effects in highly cited clinical research. The Journal of American Medical Association, 294 (2): 218-28.

[98] Ioannidis, J. P. (2005b). Why most published research findings are false. PLoS Med , 2(8): e124.

[99] JavaBayes. (2001). Bayesian Networks in Java: User manual and download. JavaBayes - version 0.346. Retrieved July 21, 2012, from http://www.cs.cmu.edu/~javabayes/Home/index.html

[100] Jensen, F. V., Olesen, K. G. and Andersen, S. K. 1990. An algebra of Bayesian belief universes for Knowledge-Based Systems. Networks 20(5), 637–659.

[101] Jensen F, Bayesian Networks and Decision Graphs, Springer-Verlag, New York, 2001

[102] Jiang X., and Cooper GF. (2010). "A Bayesian spatio-temporal method for disease outbreak detection". Journal of the American Medical Informatics Association 17 (4): 462–71.

[103] Jiang, X., Neapolitan, R.E., Barmada, M.M., and Visweswaran, S. (2011). Learning Genetic Epistasis using Bayesian Network Scoring Criteria. BMC Bioinformatics 12: 89.

[104] jNBC Toolkit. (2004). Bayesian Network Classifier Toolbox. jNBC Toolbox. Retrieved July 21, 2012, from http://jbnc.sourceforge.net/#jBNC-WEKA

[105] JNCC2. (2008). The Java Implementation of Naive Credal Classifier 2. JNCC2. Retrieved July 21, 2012, from http://www.idsia.ch/~giorgio/jncc2.html

[106] Jolliffe, I. T., and Stephenson, D. B. (2003). Forecast Verification. A Practitioner's Guide in Atmospheric Science. West Sussex, England: Wiley.

[107] Jose, V. R., Nau, R. F., and Winkler, R. L. (2009). Sensitivity to Distance and Baseline Distribution in Forecast Evaluation. Management Science, 55, 582-590.

[108] Joseph, A., Fenton, N., and Neil, M. (2006). Predicting football results using Bayesian nets and other machine learning techniques. Knowledge-Based Systems, 7, 544-553.

[109] Jullien, B., and Salanie, B. (2000). Estimating preferences under risk: the case of racetrack bettors. Journal of Political Economy, 108, 503-530.

[110] Kadane, J. B., and Schum, D. A. (1996). A Probabilistic Analysis of the Sacco and Vanzetti Evidence. New York: Wiley.

[111] Karlis, D., and Ntzoufras, I. (2000). On modelling soccer data. Student, 229–244.

[112] Karlis, D., and Ntzoufras, I. (2003). Analysis of sports data by using bivariate Poisson models. The Statistician, 52: 3, 381-393.

[113] Killeen, P.R. (2005a). An alternative to null-hypothesis significance tests. Psychological Science, 16, 345–353.

[114] Killeen, P.R. (2005b). Replicability, confidence, and priors. Psychological Science, 16, 1009–1012.

[115] Knorr-Held, L. (1997) Hierarchical Modelling of Discrete Longitudinal Data, Applications of Markov Chain Monte Carlo. Munich: Utz.

[116] Knorr-Held, L. (2000). Dynamic rating of sports teams. The Statistician, 49, 261–276.

[117] Knowledge Industries. (2006). DXpress - To build the Network. DXpress. Retrieved July 20, 2012, from http://www.kic.com/dxpress.htm

[118] Koning, R. (2000). Balance in competition in Dutch soccer. The Statistician, 49: 3, 419-431.

[119] Koning, R., Koolhaas, M., Renes, G., and Ridder, G. (2003). A simulation model for football championships. European Journal of Operational Research, 148, 268-276.

[120] Kontkanen, P., Myllymaki, P., Silander, T., and Tirri, H. (1998). BAYDA: Software for Bayesian Classification and Feature Selection. Proceedings of The Fourth International Conference on Knowledge Discovery and Data Mining (pp. 254–258). Menlo Park, CA: AAAI Press.

[121] Koumenides, C. L., and Shadbolt, N. R. (2012). Combining link and content-based information in a Bayesian inference model for entity search. In Proceedings of the 1st Joint International Workshop on Entity-Oriented and Semantic Search (JIWES '12). ACM, New York, NY, USA, Article 3, 6 pages.

[122] Krustrup, P. and Bangsbo, J. (2001). Physiological demands of top-class soccer refereeing in relation to physical capacity: effect of intense intermittent exercise training. Journal of Sports Science, 19:881-891.

[123] Krustrup, P., Mohr, M., Amstrup, T., Rysgaard, T., Johansen, J., Steensberg, A., Pedersen, P. K., and Bangsbo, J. (2003). The Yo-Yo intermittent recovery test: physiological response, reliability, and validity. Medicine and Science in Sports and Exercise, 35:697-705.

[124] Krustrup, P., Mohr, M., Nybo, L., Majgaard Jensen, J., Jung Nielsen, J., and Bangsbo, J. (2006). The Yo-Yo IR2 Test: physiological response, reliability, and application to elite soccer. Medicine and Science in Sports and Exercise, 38:1666-1673.

[125] Kuonen, D. (1996). Statistical Models for Knock-Out Soccer Tournaments. Technical Report, Department of Mathematics, Ècole Polytechnique Federale de Lausanne.

[126] Kuypers, T. (2000). Information and efficiency: an empirical study of a fixed odds betting market. Applied Economics, 32, 1353-1363.

[127] Langseth, H. and Bangsø, O. (2001). Parameter Learning in Object-Oriented Bayesian Networks. Annals of Mathematics and Artificial Intelligence, 32(1): 221-243.

[128] Lauritzen, S. L. 1995. The EM algorithm for graphical association models with missing data. Computational Statistics & Data Analysis 19(2), 191–201.

[129] Lee, A. J. (1997) Modeling scores in the Premier League: is Manchester United really the best? Chance, 10, 15–19.

[130] Leitch, G., and Tanner, J. E. (1991). Economic Forecast Evaluation: Profits Versus The Conventional Error Measures. American Economic Association , 580-590. 35

[131] Leitner, C., Zeileis, A., and Hornik, K. (2010). Forecasting sports tournaments by ratings of (prob)abilities: A comparison for the EURO 2008. International Journal of Forecasting, 26, 471-481.

[132] Levitt, S. D. (2004). Why are gambling markets organised so differently from financial markets? The Economic Journal, 114, 223-246.

[133] Loftus, G. R. (1991). On the tyranny of hypothesis testing in the social sciences. CONTEMPORARY PSYCHOLOGY , 36(2): 102-105.

[134] Logic Programming Associates. (2012). Find out more about Flint. FLINT toolkit - Details. Retrieved July 20, 2012, from http://www.lpa.co.uk/fln_det.htm

[135] Lumina Decision Systems (2012). Bringing Clarity to Difficult Decisions. Retrieved June 29, 2012, from http://www.lumina.com/

[136] Macdonald, R.R. (2005). Why replication probabilities depend on prior probability distributions: A rejoinder to Killeen (2005). Psychological Science, 16, 1007–1008.

[137] Maher, M. J. (1982). Modelling association football scores. Statististica Neerlandica, 36, 109– 118.

[138] Min, B., Kim, J., Choe, C., Eom, H., and McKay, R. B. (2008). A compound framework for sports results prediction: A football case study. Knowledge-Based Systems, 21, 551- 562.

[139] Mintel Intelligence Report. (2001). Online betting. London: Mintel International Group Ltd.

[140] Mohr, M., Ellingsgaard, H., Andersson, H., Bangsbo, J., and Krustrup, P. (2004). Physical demands in high-level female soccer -application of fitness tests to evaluate match performance. Journal of Sports Science, 22: 552.

[141] MSBNx. (2012). Microsoft Research: Microsoft Bayesian Network Editor. MSBNx: Bayesian Network Editor and ToolKit. Retrieved July 21, 2012, from http://research.microsoft.com/en-us/um/redmond/groups/adapt/msbnx/

[142] Mueller, F. O., Cantu, R. C., and Camp, S. P. (1996). Catastrophic Injuries in High School and College Sports. Champaign: Human Kinetics, p. 57.

[143] Mujika, I., Santisteban, J., Angulo, P., Padilla, S. (2007). Individualized aerobic-power training in an underperforming youth elite association football player. International Journal of Sports Physiology and Performance, 2:332-335.

[144] Mujika, I., Santisteban, J., Impellizzeri, F. M., Castagna, C. (2008). Fitness determinants of success in men's and women's football. Journal of Sports Science, 1-8, iFirst.

[145] Murali, V. (2011). Bleacher Report. Retrieved March 28, 2012, from World Football: 40 Biggest Scandals in Football History: http://bleacherreport.com/articles/909932-world-football-40-biggest-scandals-in- football-history

[146] Murphy, A. (1969). On the "ranked probability score". Journal of Applied Meteorology, 8, 988-989.

[147] Murphy, A. (1970). The ranked probability score and the probability score: A comparison. Monthly Weather Review, 98, 917-924.

[148] Murphy, A., and Winkler, R. (1987). A general framework for forecast verification. Monthly Weather Review, 115, 1330-1338.

[149] Myung, I. (2003). Tutorial on maximum likelihood estimation. Journal of Mathematical Psychology, 47, 90-100.

[150] Neapolitan R.E. (2004). Learning Bayesian networks, Pearson Prentice Hall.

[151] Neil, M., Marquez, D. And Fenton, N. E. (2010). Improved Reliability Modeling using Bayesian Networks and Dynamic Discretization. Reliability Engineering & System Safety, 95(4), 412-425.

[152] Nickerson, R.S. (2000). Null hypothesis statistical testing: A review of an old and continuing controversy. Psychological Methods, 5, 241–301.

[153] Norsys Software Corp. (1995). Netica. Norsys Software Corp. - Bayes Net Software. Retrieved July 20, 2012, from http://www.norsys.com/

[154] Palisade. (2012). New PrecisionTree® 5.7. PrecisionTree: Decision Making with Decision Trees & Influence Diagrams - Palisade Corporation. Retrieved July 20, 2012, from http://www.palisade.com/precisiontree/

[155] Paton, D., and Vaughan Williams, L. (2005). Forecasting Outcomes in Spread Betting Markets: Can Bettors Use 'Quarbs' to Beat the Book? Journal of Forecasting, 24, 139-154.

[156] Pearl J, Causality: Models, Reasoning, and Inference, Cambridge University Press, 2000.

[157] Peel, D. A., and Thomas, D. A. (1989). Outcome uncertainty and the demand for football. Scottish Journal of Political Economy, 35, 242-249.

[158] Peel, D. A., and Thomas, D. A. (1992). The demand for football: some evidence on outcome uncertainty. Empirical Economics, 4, 567-570.

[159] Peel, D. A., and Thomas, D. A. (1997). Handicaps, outcome uncertainty and attendance demand. Applied Economic Letters, 4, 567-570.

[160] pi-football (2010) Probabilistic Intelligence in Football, Retrieved May 25, 2012, from Forecasts for season 2010/11: http://www.pi-football.com/past_predictions.aspx

[161] PNL. (2010). The Open Source Probabilistic Networks Library. PNL. Retrieved July 21, 2012, from http://sourceforge.net/projects/openpnl/

[162] Pope, P., and Peel, D. (1989). Information, prices and efficiency in a fixed-odds betting market. Economica, 56, 323-341.

[163] Poulter, D. R. (2009). Home advantage and player nationality in international club football. Journal of Sports Sciences , 27(8): 797-805.

[164] Pourret, O., Naim, P., and Marcot, B. (2008). Bayesian Networks: A Practical Guide to Applications. Chichester, UK: Wiley.

[165] Pulcinella. (1996). A tool for Propagating Uncertainty through Local Computations. Pulcinella Home Page. Retrieved July 21, 2012, from http://iridia.ulb.ac.be/pulcinella/Welcome.html#bib

[166] Quandt, R. E. (1986). Betting and equilibrium. Quarterly Journal of Economics, 101, 201- 207.

[167] Ridder, G., Cramer, J., and Hopstaken, P. (1994). Estimating the Effect of a Red Card in Soccer. Journal of the American Statistical Association, 89: 427, 1124-1127.

[168] Rossett, R. N. (1971). Weak experimental verification of the expected utility hypothesis. Review of Economic studies, 38, 481-492.

[169] Rotshtein, A., Posner, M., and Rakytyanska, A. (2005). Football predictions based on a fuzzy model with genetic and neural tuning. Cybernetics and Systems Analysis, 41: 4, 619- 630.

[170] Rue, H., and Salvesen, O. (2000). Prediction and retrospective analysis of soccer matches in a league. The Statistician, 49, Part 3, pp. 339-418.

[171] Sauer, R. D. (1998). The Economics of Wagering Market. Journal of Economic Literature, 36, 2021-2064. - 22 -

[172] Sauer, R., Brajer, V., Ferris, S., and Marr, M. (1988). Hold your bets: another look at the efficiency of the gambling market for national football league games. Journal of Political Economy, 96, 206-213.

[173] Schervish, M. J. (1996). P Values: What They Are and What They Are Not. The American Statistician , 50 (3): 203–206.

[174] Scout7. (2012). Scout7 Home. Retrieved online August 30, 2012, from Scout7: About Us: http://info.scout7.com/AboutUs.aspx?lang=en

[175] Shachter, R. D. 1986. Evaluating influence diagrams. Operations Research 34(6), 871–882.

[176] Shenoy, P.P. and Shafer, G. (1988). An Axiomatic Framework for Bayesian and Belief-Function Propagation. Proceedings of the AAAI Workshop on Uncertainty in AI 307-314.

[177] Shin, H. (1991). Optimal betting odds against insider traders. Economic Journal, 101, 1179- 1185.

[178] Shin, H. (1993). Measuring the incidence of insider trading in a market for state-contingent claims. Economic Journal, 103, 1141-1153.

[179] Shin, R. E. (1992). Prices of state contingent claims with insider traders, and the favourite longshot bias. Economic Journal, 102, 426-435.

[180] Snowberg, E. (2010). Explaining the Favorite-Long Shot Bias: Is it Risk-Love or Misperceptions? Journal of Political Economy, 118, 4, 723-746.

[181] Snyder, W. N. (1978). Horse racing: testing the efficient markets model. Journal of Finance, 33, 1109-1118.

[182] Sobel, R. S., and Raines, S. T. (2003). An examination of the empirical derivatives of the favourite-longshot bias in racetrack betting. Applied Economics, 35, 371-385.

[183] Soccer Base. (2012). Soccer Base. Retrieved February 23, 2012, from Man City: Full Transfer History: http://www.soccerbase.com/teams/team.sd?team_id=1718&teamTabs=transfers

[184] Sterne, J. A. (2001). Sifting the evidence—what's wrong with significance tests? BMJ, 322 (7280): 226–231.

[185] Thaler, R., and Ziemba, W. (1988). Parimutuel betting markets: Racetracks and lotteries. Journal of Economic Perspectives, 2, 161-174.

[186] Top 100 bookmakers. (2006). Retrieved November 28, 2011, from Comparative directory of online bookmakers: http://www.top100bookmakers.com/

[187] Tsakonas, A., Dounias, G., Shtovba, S., and Vivdyuk, V. (2002). Soft computing-based result prediction of football games. The First International Conference on Inductive Modelling (ICIM'2002). Lviv, Ukraine.

[188] Uebersax, J. (2004). Genetic Counseling and Cancer Risk Modeling: An Application of Bayes Nets. Marbella, Spain: Ravenpack International.

[189] US Department of Transportation. (2008). United States Department of Transportation. Retrieved August 09, 2012, from http://www.dot.gov/

[190] Vallverdú, Jordi (2008). The False Dilemma: Bayesian vs. Frequentist. E-Logos: Electronic Journal for Philosophy, ISSN: 1211-0442

[191] Vaughn Williams, L., and Paton, D. (1997). Why is there a favourite-longshot bias in British racetrack betting markets? Economic Journal, 107, 150-158.

[192] Vecer, J., Kopriva, F., and Ichiba, T. (2009). Estimating the Effect of the Red Card in Soccer: When to Commit an Offense in Exchange for Preventing a Goal Opportunity. Journal of Quantitative Analysis in Sports, 5: Iss. 1, Article 8.

[193] Vergin, R. C., and Scriabin, M. (1978). Winning strategies for wagering on National Football League games. Management Science, 24, 809-818.

[194] Wagenmakers, E., and Grunwald, P. (2006). A Bayesian Perspective on Hypothesis Testing. A Comment on Killeen (2005). Psychological Science. Volume 17, Number 7.

[195] Webby, R., and O'Connor, M. (1996). Judgemental and statistical time series forecasting: A review of the literature. International Journal of Forecasting, 12, 91-118.

[196] Wilks, D. (1995). Statistical Methods in the Atmospheric Sciences. International Geophysics Series, Vol. 59, Academic Press, pp. 467.

[197] William Hill PLC. (2012). Retrieved January 25, 2012, from About William Hill: http://www.williamhillplc.com/wmh/about/

[198] Wing, C., Tan, K., and Yi, L. (2007). The Use of Profits as Opposed to Conventional Forecast Evaluation Criteria to Determine the Quality of Economic Forecasts. Singapore: Nanyang Business School, Nanyang Technological University, Nanyang Avenue.

[199] Woodland, L., and Woodland, B. (1994). Market efficiency and the favourite-longshot bias: the baseball betting market. The Journal of Finance, 49, 269-279.

[200] Woodland, L., and Woodland, B. (2001). Market efficiency and profitable wagering in the National Hockey League: can betting score in longshots? Southern Economic Journal, 67, 983-995.

[201] Ziliak, S. T., and McCloskey, D. N. (2004). Size Matters: The Standard Error of Regressions in the American Economic Review. Journal of Socio-Economics , 33 (5), 527-546.

[202] Ziliak, S. T., and McCloskey, D. N. (2008). The Cult of Statistical Significance: How the Standard Error Costs Us Jobs, Justice, and Lives (Economics, Cognition, and Society). The University of Michigan Press .

[203] Zuber, R., Gandar, J., and Bowers, B. (1985). Beating the spread: testing the efficiency of the gambling market for national football league games. Journal of Political Economy, 93, 800-806.