

Towards Smart-Data: Improving predictive accuracy in long-term football team performance

Anthony Constantinou^{a, b} and Norman Fenton^b

- a. Corresponding author. E-mail address: anthony@constantinou.info
- b. Risk and Information Management (RIM) Research Group, School of Electronic Engineering and Computer Science, Queen Mary University of London, London, UK, E1 4NS.

THIS IS A DRAFT OF THE ACCEPTED VERSION OF THE FOLLOWING CITATION:

Constantinou, A. C. and Fenton, N. (2017). Towards Smart-Data: Improving predictive accuracy in long-term football team performance. *Knowledge-Based Systems*, In Press, 2017.

DOI: <http://dx.doi.org/10.1016/j.knosys.2017.03.005>

Corresponding author: Dr. Anthony Constantinou, E-mail: anthony@constantinou.info

© 2017. This manuscript version is made available under the CC-BY-NC-ND 4.0 license: <http://creativecommons.org/licenses/by-nc-nd/4.0/>



ABSTRACT: Despite recent promising developments with large datasets and machine learning, the idea that automation alone can discover all key relationships between factors of interest remains a challenging task. Indeed, in many real-world domains, experts can often understand and identify key relationships that data alone may fail to discover, no matter how large the dataset. Hence, while pure machine learning provides obvious benefits, these benefits may come at a cost of accuracy. Here we focus on what we call smart-data; a method which supports data engineering and knowledge engineering approaches that put greater emphasis on applying causal knowledge and real-world ‘facts’ to the process of model development, driven by what data are really required for prediction, rather than by what data are available. We demonstrate how we exploited knowledge to develop a model that generates accurate predictions of the evolving performance of football teams based on limited data. The model enables us to predict, before a season starts, the total league points a team is expected to accumulate throughout the season. The results compare favourably against a number of other relevant and different types of models, and are on par with some other models which use far more data. The model results also provide a novel and comprehensive attribution study of the factors most influencing change in team performance, and partly address the cause of the widely accepted favourite-longshot bias observed in bookies odds.

Keywords: data engineering; dynamic Bayesian networks; expert systems; favourite-longshot bias; football predictions; knowledge engineering; smart data; soccer predictions.

1. INTRODUCTION

In many application domains, such as consumer shopping behaviour, applying state-of-the-art machine learning methods to very large datasets (often referred to as ‘big data’) can reveal new insights that would otherwise remain unknown. However, relying purely on large datasets for automated knowledge discovery and prediction can be inappropriate in many real-world domains. For example, the 2007-09 financial crisis revealed that big-data models used by investment banks and rating agencies failed to predict real-world financial risk. Wieland and Wolters (2012) demonstrate how each of the 50+ professional economic models failed to foresee the downturn. This is because such big-data models do not incorporate new and, hence, previously unseen factors - in this case the sudden burst of the USA housing ‘bubble’ due to high default rates in the subprime mortgage sector after many high risk loans were sold on as low risk securities (IMF, 2009). Similarly, it would be doubtful whether such models could accurately predict the economic consequences of a country leaving the EU.

This paper focuses on a prediction problem that has similarities to financial risk, namely predicting evolving football team performance. In both domains future performance can be suddenly and dramatically affected by rarely seen events and so both require smarter ways to perform data engineering and modelling, rather than rely on ever larger amounts of data.

Football is the world’s most popular sport and constitutes an important share of the gambling market. This has motivated numerous researchers to use football events as a real-world application domain to assess various statistical, probabilistic and machine learning techniques. The most common problem addressed is that of predicting the outcome of specific football matches. This started with the popular work of Maher (1982) on a bivariate Poisson model to predict match results based on the attack and defence capabilities from goals scored and conceded, and which was later adopted by Dixon and Coles (1997) for the purpose of generating profit against published market odds. Since then, numerous Poisson-based models have been published. Rue and Salvesen (2000) proposed a time-series Poisson distribution model and demonstrated profitability against Intertops odds; this model was later revised by Cowder et al. (2002) to a computationally less demanding model. Karlis and Ntzoufras (2003) also assessed double Poisson and bivariate Poisson models and concluded that the bivariate Poisson model correlates well between scores of competing teams.

Football predictions also became popular with computer scientists, primarily for the purpose of testing machine learning techniques. These include: Tsakonas et al. (2002) who found genetic programming to be superior to fuzzy models and neural networks in predicting football; Rotshtein et al. (2005) who concluded that the combination of genetic and neural optimisation techniques can lead to ‘acceptable’ football match predictions; Joseph et al. (2006) who showed that incorporating expert judgments with data into a Bayesian network (BN) model led to more accurate predictions compared to a number of other data-driven machine learning techniques; Min et al. (2008) who combined BNs with rule-based reasoners to predict World Cup matches; Baio & Blangiardo (2010) who tested a Bayesian hierarchical model and found it to be on par with the Karlis and Ntzoufras (2003) Poisson-based model; and Constantinou et al. (2012; 2013a) who developed a BN model, based on both data and knowledge, to predict English Premier League (EPL) matches before they start (predictions were published online at www.pi-football.com), and demonstrated profitability against all of the available market odds. Such BN models now also extend to assessing referee bias in football (Constantinou et al., 2014).

Rating systems are also popular within the football predictions literature. For example, Knor-Held (2000) proposed a rating system for football teams based on the cumulative link model

for ordered responses, where latent parameters represent the strength of each team. The ELO rating system, which was initially developed to assess the strength of chess players (Elo, 1978), was assessed by Hvattum and Arntzen (2010) and showed that even though the ratings appeared to be useful in encoding the information of past results for measuring the strength of a team, the predictions were still considerably less accurate compared to market odds. The ELO rating and the ELO-based FIFA/Coca-Cola World Ranking (FIFA, 2016) were also assessed by Leitner et al. (2010) who used them to predict tournament winners, and concluded that the accuracy of both ratings was inferior to bookmakers' odds. Constantinou and Fenton (2013b) proposed the pi-rating system, which determines the level of ability of football teams based on the relative discrepancies in scores between adversaries, and showed that it outperformed ELO-based rating systems and was also able to generate some profit against published market odds over a series of five EPL seasons. Such rating systems are now also extended to determining player skill (McHale et al., 2012).

The popularity of football models is in many cases driven by the prospect of identifying market inefficiencies in terms of the published odds available for betting. Sauer (1998) suggested that a betting market can be considered efficient only if profitable opportunities cannot be exploited. While some studies suggest that the football market may be efficient, due to either not discovering profitable betting strategies or due to market odds outperforming the predictions generated by various proposed statistical models (Peel & Thomas, 1988; 1992; 1997; Pope & Peel, 1989; Forrest et al., 2005; Graham & Stott, 2008; Vecer et al., 2009), numerous other authors have concluded that the market is inefficient. For example, many of the Poisson-based studies have claimed market inefficiency due to generating positive returns against published odds under specific conditions (Dixon and Coles, 1997; Rue and Salvesen, 2000; Kuypers, 2000; Dixon & Pope, 2004). Other studies have reported inefficiencies at the start and at the end of a football season (Goddard & Asimakopoulos, 2004), and that the odds offered on popular teams are inefficient due to being offered more favourable terms on their wagers (Forrest & Simmons, 2008). Studies which demonstrate consistent profitability throughout a football season also question the efficiency of the market (Constantinou et al., 2012; 2013a). However, the strongest evidence of inefficiency probably comes from studies that do not employ predictive models, but rather assess bookmaking odds for biases, and claim inefficiency in the presence of the favourite-longshot bias (Cain et al., 2000, Forrest & Simmons, 2001; 2002); this is where bets on favourites tend to have higher expected returns than bets on longshots. This type of bias seems to extend to home-team bias and most-likely outcome bias (Constantinou & Fenton, 2013c).

Most of the previous extensive work on results predictions has primarily used historical data of relevant match results. In this paper we do not consider individual match results, but rather exploit external factors which may influence the strength of a team and resulting performance. We are interested in predicting a football team's performance for a whole season (measured by total number of points won) before the season starts. This is an important and enormous gambling market in itself - bettors start placing bets such as which team will win the title, finish in the top four, or be relegated, as soon as the previous season ends. The need for greater accuracy in such predictions has become the subject of international interest following the 2015-16 EPL season when Leicester City finished top of the league, having been priced at 5,000 to 1 by many bookmakers before the start of the season.

While a team's most recent previous seasons' performances are clearly important, changes may occur over the summer break (such as the purchase of new players or other staff) that may have a major influence on the team's future performance. Since no competitive matches occur over the summer break, it is impossible to judge changes in team strength until the new season begins and match results are observed. Moreover, a team's most recent season performance may

not accurately reflect their true strength if, for example, the team had been adversely affected by major injuries from which players have now recovered.

While the model presented in this paper is important in terms of the application domain for the reasons highlighted above, the process by which the data are engineered and the model is developed is equally important in highlighting the benefits of a smart-data approach. The paper is organized as follows: Section 2 discusses the problem associated with data availability, Sections 3 and 4 cover the data engineering and model development processes respectively, Section 5 covers model validation, Section 6 discusses and presents further results of interest, and Section 7 provides our conclusions.

2. THE PROBLEM

It is a common practice to generate models based on what data are available. For the problem of predicting football team performance as measured by league points accumulated from season to season, there are a multitude of potential available data. However, with the exception of previous seasons' league points, there is little obviously relevant data that are readily available. The temptation in such circumstances is to find any 'easily available' data and throw regression or machine learning methods at it. Such data might include things like: average home attendances, financial turnover, number of international players, size of squad etc. In contrast, our starting point was to use knowledge to identify what data are really required to assess the evolution of team strength between seasons. Focusing on the EPL, we have identified the following factors, with the stated assumptions:

1. **League points (LP):** as discussed above.
2. **Transfer spending (T):** Player transfers (in and out) can influence team strength.
3. **Injuries (I):** Excessive injuries to key players leads to degraded team performance while fewer than expected injuries leads to improved performance.
4. **Managerial changes (M):** Different managers may employ different tactics, and may also influence team spirit and morale, amongst others.
5. **European (EU) competition involvement:** Those few teams involved in the additional European competitions (Champions League and Europa League) face greater demands than those that are not. Their performance in the league may be degraded because of shift in focus away from the league and increased player fatigue due to both the higher number of matches required to be played and the extensive travelling requirement to other countries.
6. **Promoted team (P):** If a team is new to the EPL due to promotion from the (lower) Championship division, the normal 'benchmark' performance is that of the team they replace. However, *promoted teams* often have special circumstances that make their performance relative to the team they replace very different.

In what follows we explain why a typical pure data-driven model (e.g. regression) which attempts to predict next season's league points as a function $f(LP, T, I, M, EU, P)$, of the above specified

factors is inappropriate for time-series inference in football. However, first we have to examine whether the dataset is structured correctly, in terms of adhering to this type of time-series influential modelling, otherwise it would be impossible to build an appropriate model.

2.1 IDENTIFYING MODEL REQUIREMENTS

To justify the restructuring of the data, we require a good understanding in terms of how the model is expected to simulate the real-world behaviour of the problem specified. We employ a Bayesian Network (BN) model for this purpose. Figure 1 illustrates the simplified topology of the overall BN model.

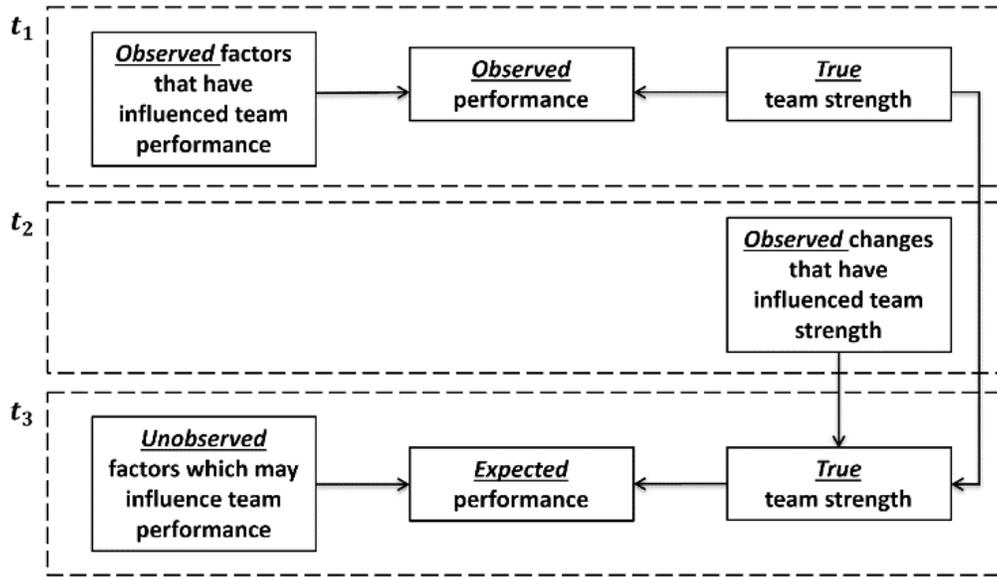


Figure 1: Simplified model topology of the BN model.

Time-series analysis requires that we take into consideration the following three time steps:

Time step t_1 : This first time step represents the previous season. The relevant factors are:

1. **Observed factors that have influenced team performance:** This node represents a BN component consisting of relevant variables such as player injuries and involvement in European competitions;
2. **Observed performance:** This node represents a single variable which takes as input the total league points accumulated over the previous season;
3. **True team strength:** This node represents the latent variable that we are interested in inferring at t_1 , based on the observations provided for nodes which exist in 1 and 2 above. In brief, the *true* strength of a team represents team performance independent of component (1) factors, such as injuries. If we observe performance X points, in conjunction with factors that are expected to have resulted in changing team performance by Y points (Y may be positive or negative), then the *true* strength of that team is $X - Y$

points. This is achieved by setting the Conditional Probability Table (CPT) of node X to $\{Y + \text{true strength}\}$.

Time step t_2 : This second time step represents events of team changes occurring over the summer break. The model assumes that summer breaks occur at the end of the previous season (i.e., t_1) and prior to the start of the next season (i.e., t_3). At t_2 , the node **Observed factors that have influenced team strength** is another BN component which consists of relevant variables that represent events which may occur over the summer break, such as player transfers.

Time step t_3 : The third and final time step, t_3 , represents the subsequent season which commences following the summer break. The topology of this time step consists of:

1. **Unobserved factors which may influence team performance:** This is a BN component identical to that of time step t_1 , with the only difference being that the factors which may influence team performance are unobserved (since this component represents the subsequent season under assessment);
2. **True team strength:** As in t_1 , this node represents a latent variable that we would like to infer. However, factors that influence team performance are unobserved at t_3 and hence, the *true* strength of the team is revised from t_1 to t_3 based on observations provided at t_2 .
3. **Expected performance:** This node represents the ultimate latent variable we would like to assess. This variable is identical to that of time step t_1 (i.e., *Observed performance*). At t_3 , however, team performance is unknown (hence *Expected performance*) and is inferred by conditioning on the revised *true* team strength in conjunction with the uncertainties that arise as a result of the unobserved factors which may influence team performance.

3. DATA ENGINEERING

Figure 2 illustrates how the relevant data observations (left), collected over seasons 2000/01 to 2014/15, are reconstructed (right) such that they adhere to the modelling concept of Figure 1. The subsections that follow cover the process of data engineering with respect to each of the factors identified in Section 2.

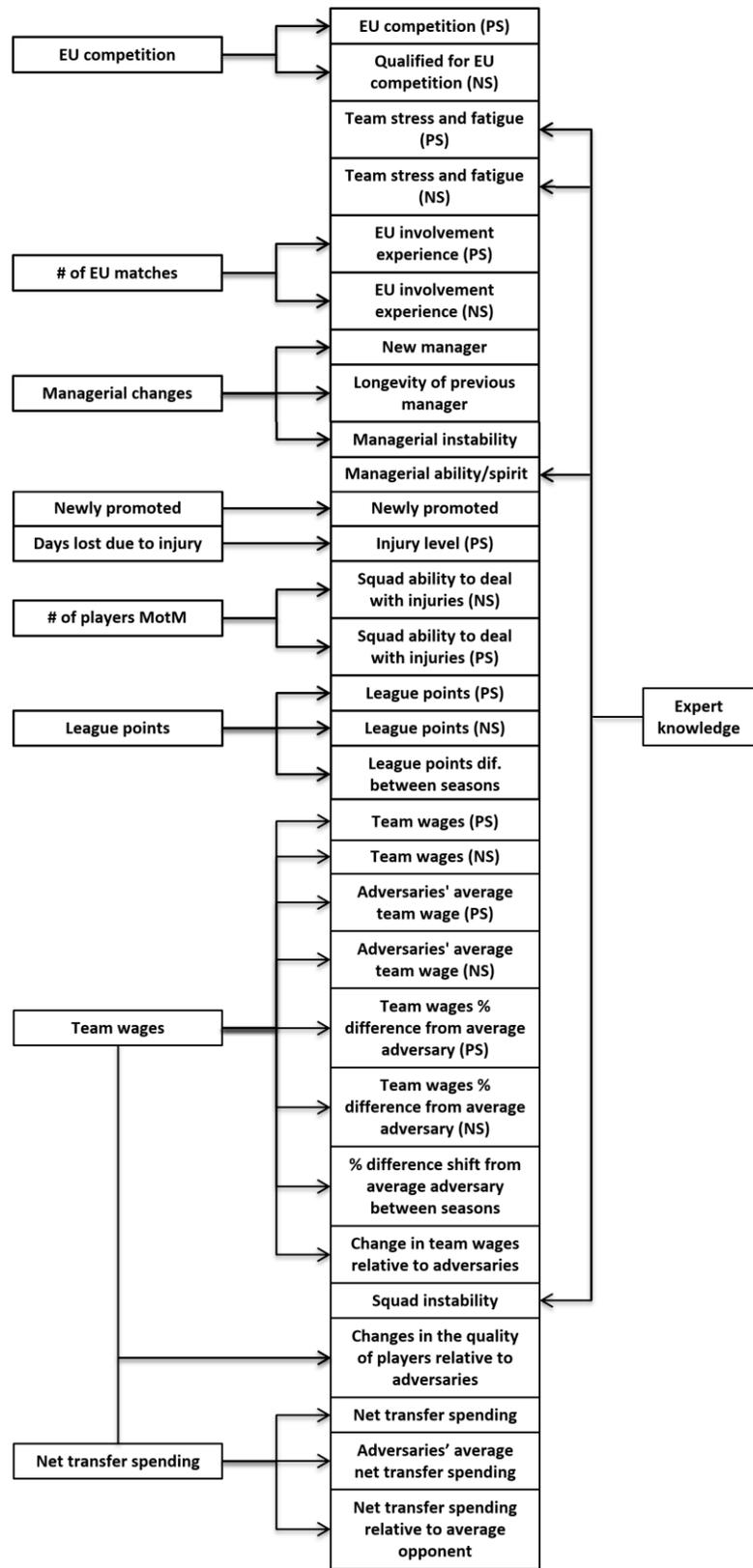


Figure 2: Restructuring the data collected (left) to that required by the model (right), where *PS* (i.e. previous season) is a variable that has already been observed during the previous season, and *NS* (i.e. next season) is a variable being assessed for the next season.

3.1 LEAGUE POINTS

Data were collected from XScores¹ (2016) and represent the total league points accumulated by each team per season. As illustrated in Figure 2, the data were split into three variables. Specifically, in terms of *training* the model for seasons s and $s + 1$, the three data variables would be as follows:

1. Points accumulated over season s (t_1);
2. Points accumulated over season $s + 1$ (t_3);
3. Discrepancy between observations (1) and (2).

3.2 INJURIES

Data were collected from PhysioRoom (2016), and represent the total number of days with injury over all players for each team per season. While this variable is duplicated into the BN (one for previous and another for next season), it was sufficient to retain one instance of this variable in the dataset based on the assumption that injuries are random.

However, it is accepted from knowledge that some teams can deal with injuries better than other teams. In implementing this belief, we found the *Man-of-the-Match* (MotM) data² provided by WhoScored (2016) to be a good indicator under the assumption that a team's ability to deal with injuries depends on the number of players who can 'make the difference' in a match. Figure 2 shows that the *MotM* variable is duplicated into the dataset. This is because we assume that the number of *MotM* players is influenced by player transfers which occur over the summer break; hence we require *MotM* observations for both s and $s + 1$, with observations at $s+1$ conditioning on player transfers that occurred at t_2 .

3.3 EUROPEAN COMPETITIONS

Data with respect to the number of European competition matches, as well as the type of competition, have been recorded manually from Wikipedia. A distinct Wikipedia page exists for each team per season. For example, in the case of Arsenal during season 2014/15, the data can be found at (Wikipedia, 2016b). Both variables are duplicated in the dataset in the same way *MotM* is, since EU qualification at $s + 1$ becomes known at the end of s and hence, influences the number of EU matches expected to be played at $s + 1$.

3.4 MANAGERIAL CHANGES

Relevant data was manually recorded from MyFootballFacts (2016). This involves data relating to changes in management which occurred both during³ a football season as well as over the summer break period. The data also capture the number of seasons the previous manager had spent at the club prior to being replaced. We did not duplicate this variable within the dataset under the

¹ We discovered that data provided by XScores (2016) had errors with respect to the EPL season 2000/01. We have cross validated the data with other sources and confirmed that the data provided by Wikipedia (2016a) were the correct data.

² Only players with total appearances greater than the average number of appearances in the EPL are considered.

³ If a team starts the season without a managerial change but experiences managerial changes some time later within the same season, we assume that such a team experienced a managerial change for the given season only if the new manager took over for at least half of the season.

assumption that most important managerial changes take effect at the end of a football season and hence we considered the variable to be a t_2 (i.e., summer break) factor.

3.5 PROMOTED TEAMS

Data were also collected from XScores (2016). This is another t_2 factor since relegations and promotions of teams between leagues occur at the end of each season. As a result, there was no need to duplicate this variable within the dataset, and no further reconstruction was deemed necessary. As a *prior* belief, we assume that the strength of the three teams promoted to the EPL is equivalent to the team strength of the three respective teams relegated to the Championship (i.e., the 2nd football division in England).

3.6 TRANSFERS

We have made use of relevant financial data, namely *Net transfer spending* and *Total team wages*, as provided by Telegraph (2016) for seasons 2000/01 to 2010/11, by TSM Plug (2016) for seasons 2011/12 to 2013/14, and by Total Sportek (2016) for season 2014/15. However, in their current structure, these data provide very limited information. This is because the current data structure fails to take into consideration the following two important factors:

1. *Relative additional spend*: If a team has a positive net investment of X million British pounds (GBPs) on player transfer spend for the upcoming season, then such a team's performance is expected to improve over the next season. If, however, every other team also has a similar positive net investment on new players, then any positive effect is diminished or cancelled.
2. *Inflation of salaries and player values*. Investing £100m to buy players during season 2014/15 is not equivalent to investing £100m to buy players during season 2000/01. The same applies to the wage increase of players over the years due to inflation.

As a result, we had to ensure that the model takes into consideration financial data in relative, rather than absolute, monetary terms. In the case of net transfer spending, we addressed this problem by simply considering the relative percentage difference in spending between teams. In the case of team wages, however, there is an extra layer of complexity. Consider team A with previous season's team wages at £100m, increasing to £120m for the subsequent season. To determine the effect of A's wage increase of 20% we need to perform comparisons against the average adversary. Suppose the wage bill of the average adversary was £50m during the previous season, and increased to £70m for the subsequent season. While the average adversary will also be spending £20m more on team wages, just like team A, this increase constitutes 40% relative to the previous season for the average adversary, and 20% for team A. Team A was spending 100% more on team wages relative to the average adversary over the previous season (i.e., £100m against £50m), whereas for the subsequent season the relative difference drops down to 71.43% (i.e., £120m against £70m). Under such a scenario, the model will consider that team A will experience a loss in terms of strength over the subsequent season, relative to the average adversary.

4. THE TIME-SERIES MODEL

In this section we demonstrate each of the BN model fragments corresponding to each of the time steps of Figure 1, as well as the process by which the BN model is used as a Dynamic BN (DBN) to assess the evolution of team strength over multiple seasons. Dynamic BNs are BNs which relate variables to each other over time steps, where both the parameters and the structure are fixed. A more general approach where you have BNs with time-indexed random variables is called Temporal BNs. We make use of this approach to model the time steps t_1 , t_2 and t_3 , and to revise them between seasons as follows (refer to Figure 1):

1. The unobserved factors at time t_3 of BN instance n become observed at time t_1 for BN instance $n + 1$;
2. The inferred distribution *Expected performance* at time step t_3 of BN instance n becomes observed at time step t_1 of BN instance $n + 1$.

The BN variables consist of both discrete and continuous *Gaussian* distributions. Team strength and resulting team performance are both measured by the total league points accumulated. All of the data factors that influence performance, such as level of injuries and changes in the quality of players, are represented by discrete variables, whereas all of the variables that measure performance or impact/change in team strength are *Gaussian*. We have made use of the AgenaRisk toolset (Agena, 2016) to build the model. AgenaRisk makes use of the dynamic discretization algorithm (Neil et al., 2007) to simulate continuous distributions without any restrictions on the type of descendant and ancestor nodes. The CPTs of the data variables are directly learned from data with no missing values. The latent variables (e.g., true team strength, expected performance) are composite variables as defined in Section 2.1.

4.1 EXPERT KNOWLEDGE

A limited number of expert variables are introduced into the model. These variables are:

1. ***Team stress and fatigue***: allows the expert to specify the degree of fatigue and/or stress caused to a team as a result of participating in additional competitions in Europe, as shown in Figure 4 of Section 4.2;
2. ***Squad instability***: allows the expert to specify the degree of instability caused to a team as a result of having too many player changes in the first-team squad due to transfers, as shown in Figure 6 of Section 4.3;
3. ***Managerial ability***: allows the expert to specify whether a team's managerial ability has increased, decreased, or remained stable, following the arrival of a new manager, as also shown in Figure 6 of Section 4.3.

The expert variables are incorporated into the model in such a way that they do not influence the data-driven expectations, as long as they remain unobserved. As an example, Figure 3 presents the part of the BN model where the expert node *Squad instability* is incorporated, and shows how the data-driven prior of the ancestor node *Impact of changes in players* is preserved⁴, as long as the expert node remains unobserved. This outcome is achieved using the technique described in (Constantinou et al., 2016), which is based on the notion that the statistical outcomes (e.g. *Impact [on team strength] of changes in players*) are already influenced by the causes an expert might identify as variables missing from the dataset (e.g. *Squad instability*); hence, the prior expectation of a data-driven node is not amended when incorporating an expert-driven node, as long as the expert variable remains unobserved within the model.

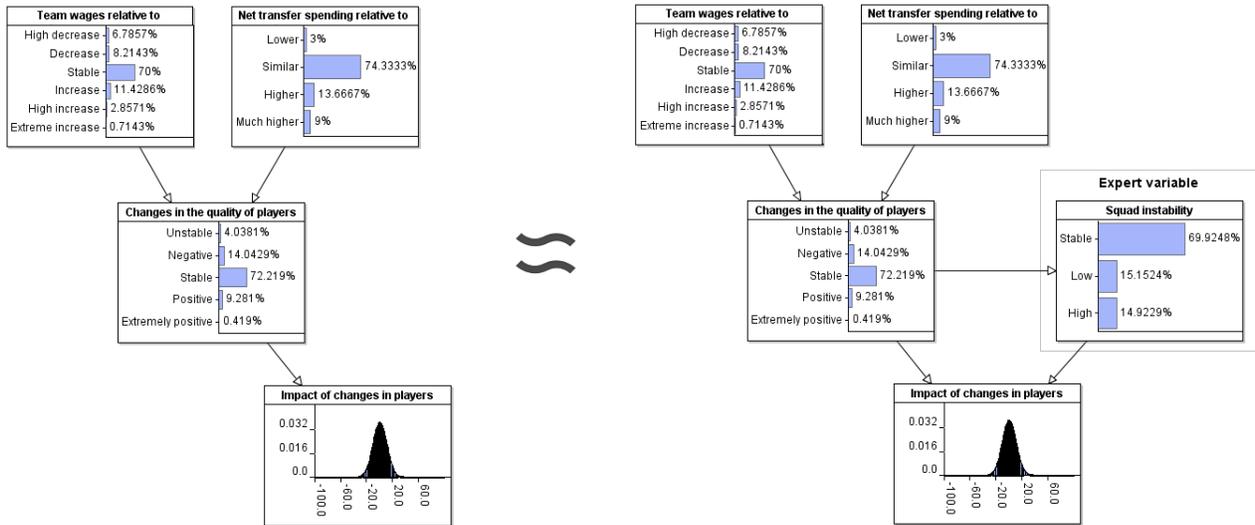


Figure 3. Extending the data-driven model (left) by incorporating the expert-driven node *Squad instability*, with the aim of preserving the data-driven prior of the ancestor node *Impact of changes in players* (right model) using the technique described in (Constantinou et al., 2016).

4.2 TIME STEPS t_1 AND t_3

Time steps t_1 and t_3 are covered jointly in this subsection since they are almost identical in terms of structure and variables considered. Figure 4 presents the BN model fragment which simulates the events at t_1 . As discussed in Section 2.1, the observed team performance, in conjunction with factors that may have influenced team performance, are taken into consideration for inference of the *true* team strength. Since the observed performance is simply the total number of points accumulated, an integer input is provided as observation for *Observed performance*. The observed factors that may have influenced team strength are assessed as follows:

1. ***Impact of European (EU) competition involvement on team performance***: This is a mixture of *Gaussian* distributions conditioned on the type of EU competition (Champions League or Europa League), the level of readiness for EU involvement (e.g. how familiar the team is in terms of participating in EU competitions), and the expert variable *Team stress and fatigue* (as discussed in Section 4.1).

⁴ Note that the technique preserves the expected value of the distribution. The shape and/or the variance of the distribution are subject to amendments (Constantinou et al., 2016).

2. **Impact of injuries on team performance:** This is a mixture of *Gaussian* distributions conditioned on the level of injuries in conjunction with the team's ability to deal with injuries (as discussed in Section 3.2).

Figure 5 presents the part of the BN model fragment that simulates the events at t_3 . The difference between components t_1 and t_3 (i.e., Figure 4 and 5) is that the factors at t_3 remain unobserved, and the prior probabilities of the variables *EU involvement readiness*, *EU competition*, and *Team ability to deal with injuries* are conditioned on observations provided to relevant factors at t_1 and t_2 as shown in Figure 5.

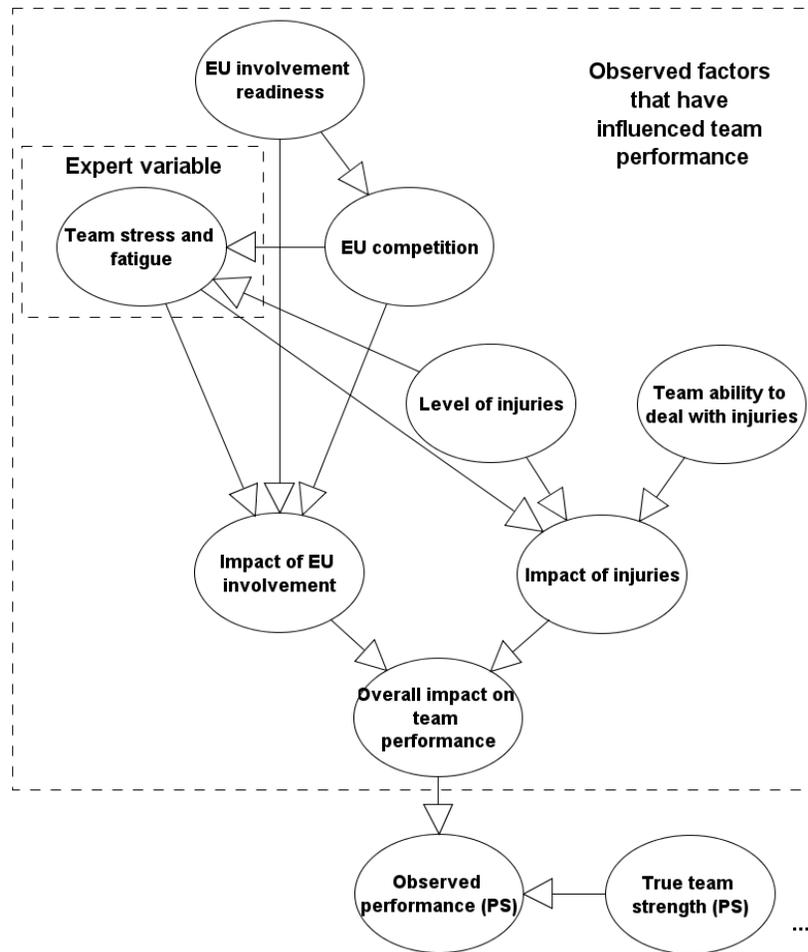


Figure 4: BN fragment t_1 , where *PS* is previous season.

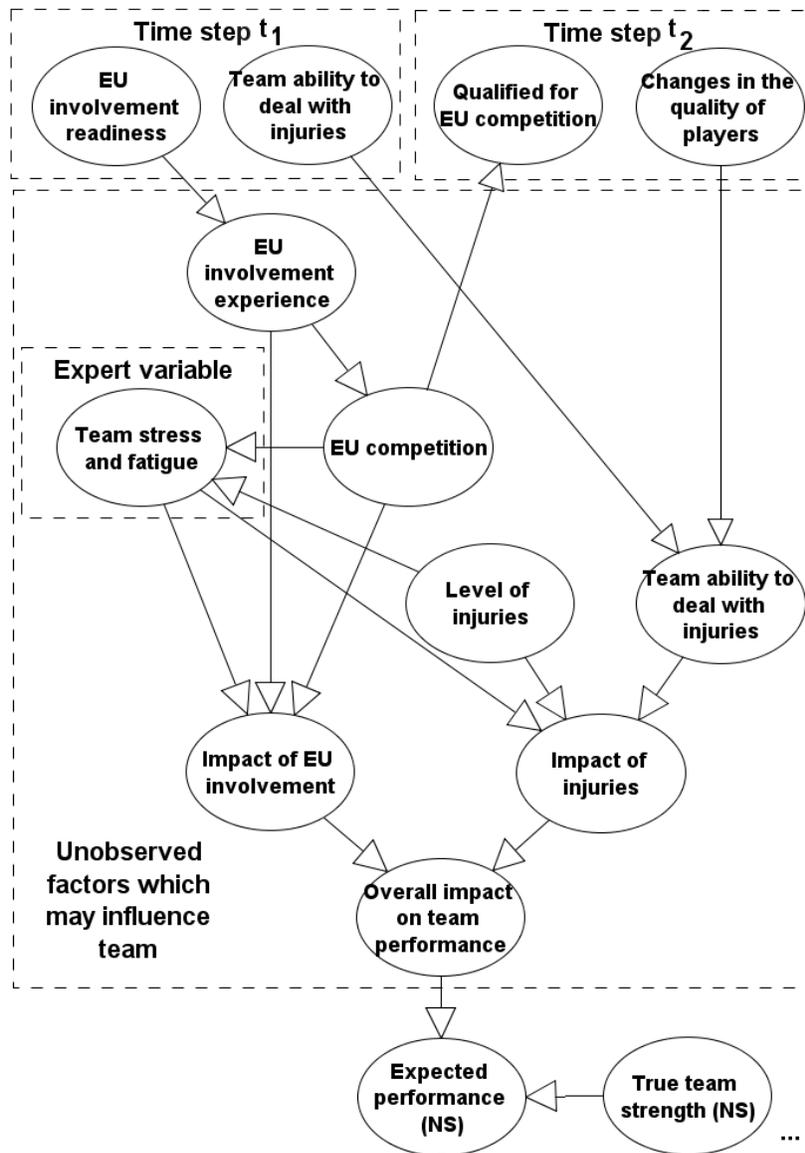


Figure 5: BN fragment t_3 , where *NS* is next season (and, at that time step, unobserved).

4.3 TIME STEP t_2

Figure 6 presents the BN model fragment that simulates the events at t_2 . The factors considered at this time step are as follows:

1. **Impact of changes in players:** Relative changes in team wages and net transfer spending are taken into consideration to formulate the composite variable *Changes in the quality of players* (relative to adversaries) which, in conjunction with the expert variable *Squad instability*, is taken into consideration in revising team strength.
2. **Impact of changes in management:** Similarly, any arrival of a new manager, in conjunction with the longevity of the departing manager, is taken into consideration to formulate the composite variable *Managerial instability*. The model assumes that the

managerial instability is conditioned on the longevity of the departing manager. For example, Sir Alex Ferguson spent 26 years as a manager for Manchester United. The replacement of such a manager affects instability in a different way when compared to the replacement of a manager who had only spent a single season at a club. Whether the new manager is believed to be superior to the departing manager can be judged expertly through the expert variable *Managerial ability*.

3. **Impact of promoted teams:** Lastly, the impact of promoted teams is solely measured based on the variable *Newly promoted*. This variable classifies the teams between promoted and non-promoted. Non-promoted teams are further classified into levels of team strength, under the assumption that the arrival of newly promoted teams may influence the performance of existing EPL teams in different ways, based on their strength.

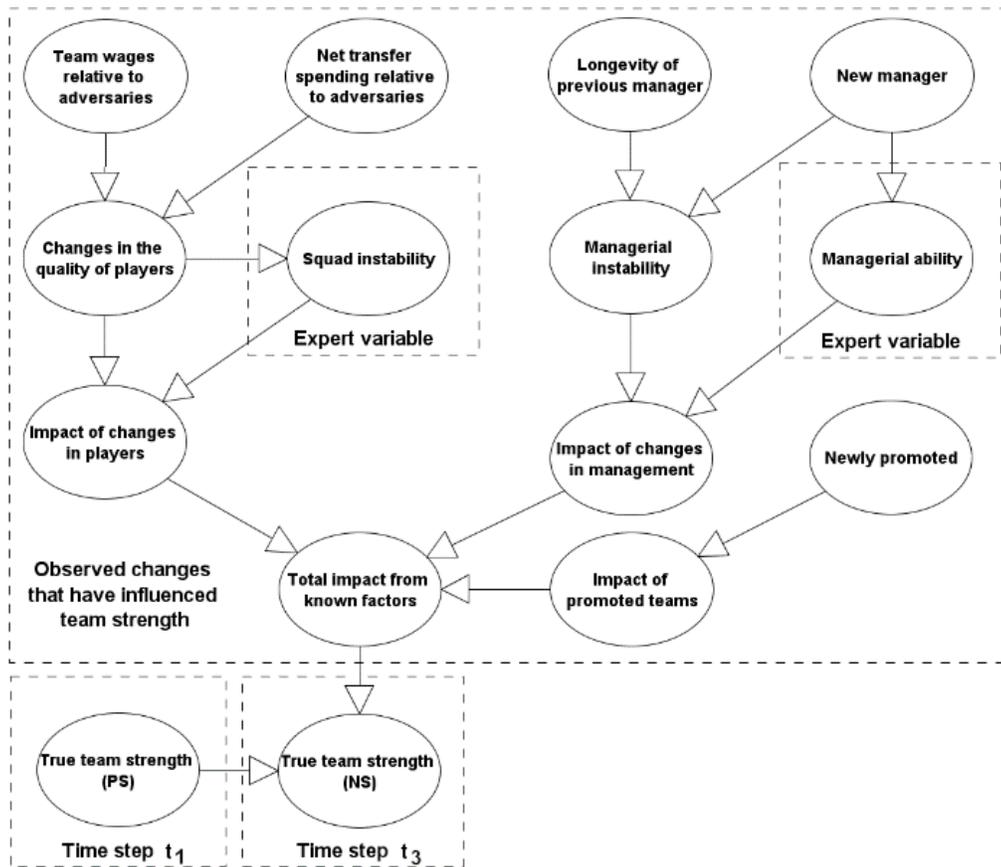


Figure 6: BN fragment t_2 , where *PS* is previous season and *NS* is next season.

5. MODEL VALIDATION

The validation is divided into two categories:

1. In Section 5.1 the model is validated in terms of how accurately it can predict team performance for each subsequent season, i.e., by assessing the predicted performance at time step t_3 for each BN instance n against the observed performance at time step t_1 for each subsequent BN instance $n + 1$.

2. In Section 5.2 the model is validated in terms of how accurately it can predict the performance of specific teams which demonstrate the highest volatility in team performance, over a series of seasons.

No similar model exists in the literature to serve as a ‘true’ comparison for model validation purposes. To assess the performance of the BN we consider three types of models or predictors. First, we consider a naïve method that serves as a lower benchmark predictor as follows:

1. **NM (Basic method)**: simply predicts the league points a team will accumulate at season $s + 1$ as the number of league points the team accumulated at season s .

Second, we have formulated two regression based models that use the dataset we collated for this study. These are:

2. **R1**: a standard non-linear regression model that considers the factors presented on the left of Figure 2 (i.e., prior to data engineering), in addition to the league points accumulated at season $s - 1$, as predictors for league points accumulated at season s . Hence, this is a regression-based predictor before any data engineering practises are performed.
3. **R2**: a second standard non-linear regression model that considers the same factors as those considered by **R1**, but in relative terms just like in the BN (e.g., refer to Section 3.6). As a result, **R2** predicts the change in league points accumulated between seasons, rather than total league points accumulated per season.

Third, we consider other published match prediction models. Validation is achieved by taking into consideration match result predictions from these models to compute the expected number of league points for each team, per season. This requires access to the predictions generated by such models, but limits the selection to models which focus on the EPL and for seasons that fall within the periods of 2000/01 and 2014/15. We have access to the predictions generated by three models. Additionally, we have also considered the bookmakers’ match odds, for a total of four match prediction models for validation purposes. These are:

4. ***pi-football v1***: A BN model used to predict the match outcomes of the EPL season 2010/11 (Constantinou et al., 2012). This model generates match predictions by considering four generic factors for each team, namely a) strength, b) form, c) psychology, and d) fatigue. The model has components corresponding to each of the four generic factors. Component (a) is determined based on relevant historical match data with a limited memory process, which ensures that the more recent match results are more important than the less recent match results. On the other hand, components (b), (c) and (d) are predominantly dependent on subjective information, such as information about the availability of primary key players, team fatigue, team spirit and team match motivation.
5. ***pi-football v2***: A BN model that was based on the philosophy of the *pi-football* model described above (hence we call them *pi-football v1* and *pi-football v2*) and which was used to predict the match outcomes of the EPL season 2011/12 (Constantinou et al., 2013a).

This model combined data with knowledge about factors such as those considered in *pi-football v1*, but extended to factors such as the toughness of previous match, days gap since previous match, EU competition and National team involvement, and head-to-head bias. The model components were structured hierarchically and this made the model less computationally demanding compared to *pi-football v1*, and demonstrated even higher profitable returns when assessed against the published market odds.

6. ***pi-ratings***: A rating system developed for the purpose to determine the level of ability of football teams (Constantinou & Fenton, 2013b). The rating system was assessed by generating match predictions for EPL seasons 2007/08 to 2011/12. This is a very simple rating system that considers goal scoring discrepancies in matches to determine team ability over time. The ratings were assessed for predictive accuracy over five EPL seasons, from 2007/08 to 2011/12, and demonstrated superior performance with respect to the widely accepted ELO-based football ratings, and the performance was also on par (or marginally superior) to bookmakers odds.
7. ***Bookmakers***: We have made use of the published market odds provided by Football-Data (2016) as predictions⁵ for the match instances of EPL seasons 2000/01 to 2014/15. We have considered the average bookmakers' odds for seasons 2005/06 to 2014/15, and the William Hill odds⁶ for seasons 2000/01 to 2004/05.

5.1 VALIDATION OF TEAM PERFORMANCE PREDICTION

We have performed *Leave-one-out cross-validation* (LOOCV) to assess the accuracy of the BN model. Since the dataset consists of the EPL seasons 2000/01 to 2014/15 inclusive, this makes a total of 300 test cases (i.e., 20 teams per season and over 15 seasons). The *LOOCV* implies that, for each validated case, the model parameters are trained on the residual 299 cases with the resulting model used to predict the 300th, previously unseen, case. This process is repeated 300 times. The results are then compared against the seven models discussed above.

However, as discussed above, only four out of the seven models considered (i.e., *NM*, *R1*, *R2*, and *Bookmakers*) have predictions generated for all of the 15 seasons, 300 cases. The *pi-ratings* predictions are restricted to five seasons (i.e., 100 cases), whereas the *pi-football v1* and *pi-football v2* predictions are restricted to a single season (i.e., 20 cases). To keep the comparison fair in these cases, we have also restricted the predictions of the BN model to the respective seasons associated with each of the three models. Consequently, the results presented in Table 1 are separated into four different sets. The main outcomes are:

- *NM* is by far the worst predictor, as expected.
- The BN smart-data approach leads to superior predictions relative to the regression models *R1* and *R2*, which are also employed with the dataset collated in this study. The regression

⁵ Match odds are translated into probabilities and then normalised so that match outcome probabilities sum up to 1 (i.e., by eliminating the profit margin).

⁶ Average bookmaking odds are not available for seasons 2000/01 to 2004/05. William Hill was the UK's biggest high street bookmaker for that period. Regardless, Constantinou et al., (2013c) showed that the normalised probabilities derived from different bookmakers' odds are almost identical.

models *do* consider injury and EU data for the season under assessment; implying that the regression models know the level of injury, the ability to deal with injuries, the type of EU competition, as well as the number of EU matches played. Conversely, these factors remain unknown for the season under prediction (i.e., t_3) in the BN model. If we were to remove these factors from the regression models we would have given an unfair advantage to the BN model on the basis that the BN does take this information into consideration at time t_1 , but only for the purpose of assessing the *true* strength of the team at t_1 . Instead, we had to give the advantage to the regression models. Still, the results suggest that the BN model gains a stronger advantage, over the regression models, from expert knowledge relating to the structure of the network and expert factors incorporated into the structure.

- *Bookmakers*, *pi-football v1*, *pi-football v2*, and *pi-ratings*, are all based on match data. This implies that these predictors benefit from model re-training after each additional match instance is observed throughout the season. This gives them an advantage with respect to the *BN*, *R1*, *R2*, and *NM* predictions which are generated before the season commences. Still, the results in Table 1 show that the BN model outperformed two out of the four match prediction models, including *Bookmakers*, and this result highlights the potential of smart-data approaches to prediction. These results are also consistent with the predictive accuracy and profitability results reported in (Constantinou et al., 2012; 2013a), even though in those studies the predictive validation was more appropriately based on a match-by-match basis, rather than expected total league points over a season.

Table 1: Average error E , along with standard error of the mean (SEM) for each model, in terms of discrepancy between predicted and observed league points accumulated per team per season, over the specified league seasons (LS). The range of league points in the EPL is 0 to 114.

M	LS	E	SEM
<i>NM</i>	2000/01 to 2014/15	8.51	± 0.3802
<i>R1</i>	2000/01 to 2014/15	7.27	± 0.7957
<i>R2</i>	2000/01 to 2014/15	7.3	± 0.3301
<i>Bookmakers</i>	2000/01 to 2014/15	5.33	± 0.2225
<i>BN</i>	2000/01 to 2014/15	4.98	± 0.2498
<i>pi-football v1</i>	2010/11	2.96	± 0.4007
<i>BN(a)</i>	2010/11	3.19	± 0.5104
<i>pi-football v2</i>	2011/12	4.83	± 0.6152
<i>BN(b)</i>	2011/12	5.77	± 0.9864
<i>pi-ratings</i>	2007/08 to 2011/12	6.13	± 0.4331
<i>BN(c)</i>	2007/08 to 2011/12	5.20	± 0.4126

Figure 7 demonstrates how the predictive error is distributed from top performing teams (ranked at the top of each graph), in terms of total points accumulated between seasons, to bottom performing teams. There are eight graphs corresponding to the BN and each of the other seven models, ordered by worse to best performance based on Table 1. A paired-moving average is superimposed on the graphs.

Almost all graphs demonstrate a consistent pattern whereby the performance of top performing teams is underestimated and vice-versa for the bottom performing teams. This is a very interesting result since this pattern serves as one of the reasons for the existence of the so-called *favourite-longshot bias*, whereby bookmakers are believed to under-price longshots and over-price favourites on purpose to maximise profit, under the assumption that a higher volume of bets are placed on longshots (Cain et al., 2000, Forrest & Simmons, 2001; 2002; Constantinou & Fenton, 2013c). However, our results show that this pattern holds for different types of models, not just for *Bookmakers*, and arrives naturally irrespective of whether expert knowledge is incorporated with data.

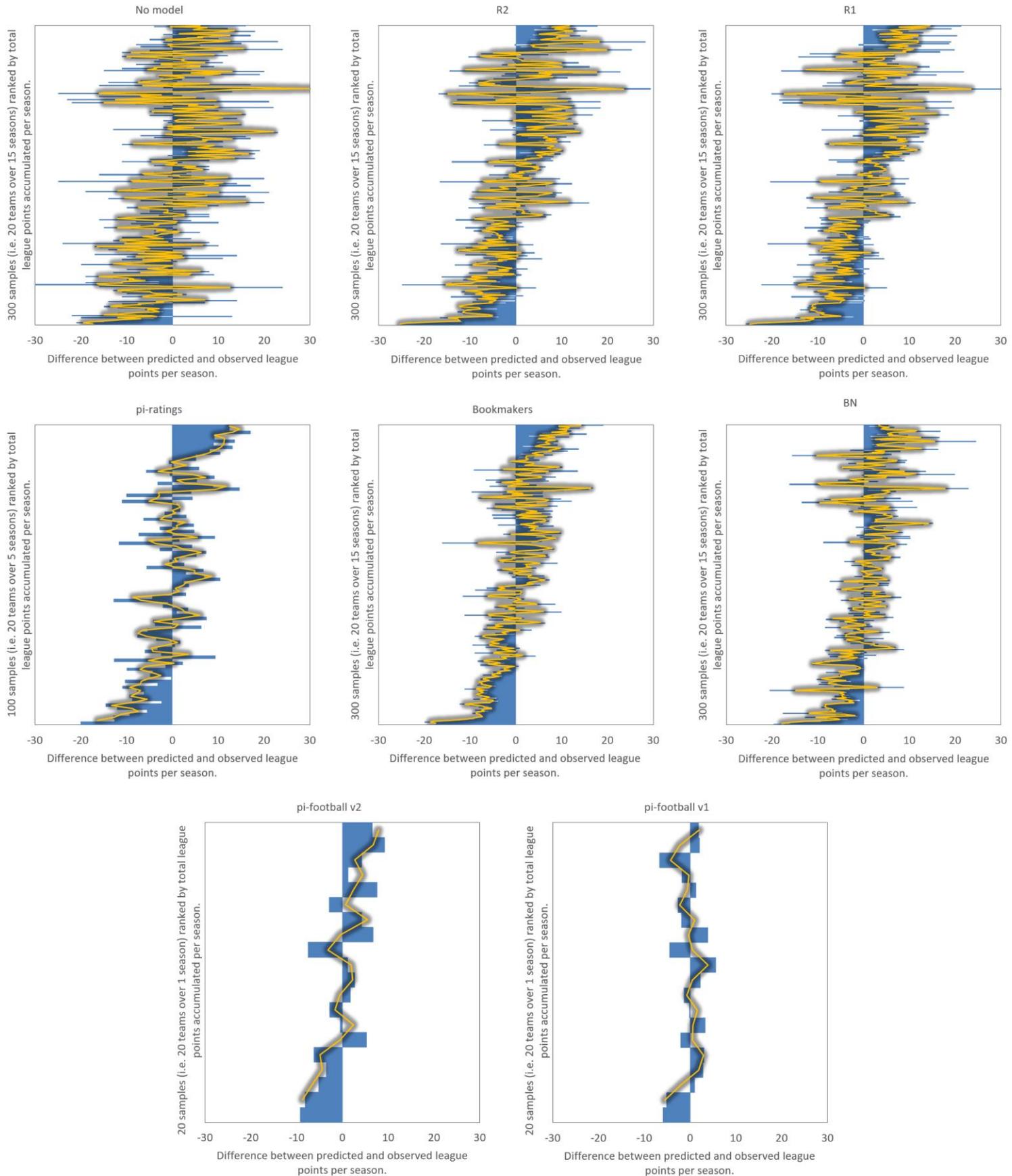


Figure 7: Predictive error (observed minus predicted) distributed from top performing teams (ranked at the top of each graph) to bottom performing teams. A paired-moving average is superimposed on the graphs.

5.2 VALIDATION OF TEAM-SPECIFIC UNSTABLE PERFORMANCES

Time-series forecasting allows us to examine whether the model is equally consistent in predicting the evolution of team strength over multiple seasons. In doing so, we have considered the teams with the most unstable performances between seasons (i.e. teams which generate high discrepancy in points between seasons). We have restricted our selection criteria to teams which participated in at least 10 out of the 15 season considered, and to teams which experience an average discrepancy between seasons of at least 9 points (i.e. equivalent to three wins/losses).

Apart from the BN model, only four out of the seven competing models considered for validation purposes satisfy these criteria. Moreover, six teams meet the strong unstable performance criterion. Table 2 presents how each of the models performs when conditioned on each of these unstable teams. The results show that the error generated by the BN model, E_{BN} , increases on average by 14.06%, which is rather faster than *Bookmakers* (i.e. $E_{Book.}$) that demonstrates an increase in error of 11.44%. On the other hand, when it comes to the other three models, the error increases noticeably faster. The results in this section are in agreement to those presented in Section 5.1, and provide evidence that the predictive performance of the BN remains rather consistent, on average, even for unstable cases. However, on an individual team basis the error can still deviate considerably from the average error demonstrated over multiple seasons. For instance, the maximum error observed by the BN for a single team in a single season is 24.57 points for Liverpool during season 2013/14, followed by the error of 22.92 points for Ipswich during season 2000/01.

Table 2: Time-series validation, where S is the number of seasons a team participated (out of 15 taken into consideration), and E_{NM} , E_{R1} , E_{R2} , $E_{Book.}$, and E_{BN} are the prediction errors generated for *No model*, *R1*, *R2*, *Bookmakers*, and the BN model respectively.

Team	S	E_{NM}	E_{R1}	E_{R2}	$E_{Book.}$	E_{BN}
Liverpool	15	11.53	9.24	10.67	7.08	6.75
Newcastle	14	11.64	10.65	9.22	7.16	5.07
Blackburn	11	11.55	6.60	8.14	5.28	6.11
West Ham	12	11.17	7.01	8.03	5.23	5.03
Everton	15	9.8	9.34	9.66	6.13	5.03
Manchester City	14	9.43	8.41	7.05	4.44	6.09
Average	-	10.81	8.73	8.69	5.94	5.68
Error increase (points)	-	2.3	1.46	1.39	0.61	0.70
Error increase (%)	-	27.03%	20.08%	19.04%	11.44%	14.06%

6. SUPPLEMENTARY RESULTS AND DISCUSSION

The results presented in this section involve the impact of specific factors of interest on team performance, in terms of league points accumulated per team, and hence provide a novel attribution study. Table 3 ranks these factors by positive influence on team performance. The results are as follows:

1. **Transfers and team wages:** Unsurprisingly, a *Much higher* net transfer spending (i.e. at least four times higher relative to the average adversary), in conjunction with an *Extreme increase* in team wages (i.e. at least 50% increase relative to the average adversary) generates the highest possible gain of 8.49 points. Rather surprising, however, is the reverse effect where *Lower* net transfer spending (i.e. at least two times lower than the average adversary) in conjunction with *High decrease* in team wages (i.e. at least 30% decrease relative to the average adversary) leads to a loss of just 1.75 points. Additionally, instability in terms of changes in players (i.e. when a team generates a high net transfer spending but team wages are reduced considerably; or vice versa) seems to generate a rather unsurprising decrease in performance of 3.73 points.
2. **EU competition:** Something that has been assumed all along, but for which no suitable analysis exists, is that the involvement in EU competitions negatively influences team performance in the league. Interestingly, the highest possible loss of 16.52 points is generated when a team participates in both⁷ EU competitions with *No/Low* EU competition readiness (i.e. previous season's EU involvement not surpassing four matches). The Europa League⁸ itself seems to carry a similar effect and generates a loss of 8.47 points. Another interesting observation is that a team with *High* EU readiness which fails to qualify for an EU competition experiences an important increase in league performance of 5.17 points. The Champions League, on the other hand, does not appear to follow the same pattern of impact. In fact, teams entering the Champions League with *No/Low* EU readiness seem to generate an increase in league performance of 1.52 points. It should be noted that Champions League qualification brings an immediate major financial reward, much of which is invested in player transfers.
3. **Managerial changes:** Changes in management do not appear to have as much of an impact on team performance as other factors do. Specifically, when a new manager joins a club, in conjunction with the departing manager's longevity with the club being *Low* (i.e. maximum of two seasons), a team's performance declines by 1.05 points. Interestingly, when the longevity of the departing manager is *High* (i.e. five seasons or more) the loss in performance is 4.64 points. On the other hand, retaining the same manager, which typically implies that the team under assessment is not under-performing (hence not seeking to replace the manager), generates an average of 0.92 points increase in performance.
4. **Injuries:** As expected, a high injury level in conjunction with the inability to deal with injuries (i.e. not many players capable of receiving the *MotM* award) leads to a significant loss of 8.31 points. On the other hand, teams who experience low levels of injury in conjunction with high ability to deal with injuries (i.e. many different players capable of receiving the *MotM* award) leads to 2.81 points increase in performance.

⁷ This can happen when a team qualifies for the Champions league and subsequently joins the Europa league due to an early disqualification from the Champions league.

⁸ Previously known as the UEFA Cup.

5. **Promoted teams:** The average promoted team generates a staggering increase in performance of 8.34 points, relative to the team being demoted in its place. This implies that the top teams of the Championship division are noticeably superior to the bottom teams of the EPL division. This outcome implies that the residual teams within the EPL are expected to lose some points, on average. The model indicates, that only the teams of *Medium* (i.e. between 50 and 69 points) and *High* (i.e. 70+ points) strength experience a decrease in performance of 3.02 and 2.91 points respectively, whereas teams of *Low* strength (i.e. less than 50 points) experience a gain in performance of 0.4 points.

Table 3: Model factors of interest and their impact on team performance, where P is the expected impact on league points for the subsequent season.

Factor/s	P
P(Net transfer spending...="Much higher"), and P(Team wages...="Extreme increase")	+8.49
P(Newly promoted="Yes")	+8.34
P(EU competition="No"), and P(EU readiness="High")	+5.17
P(Injury level="Low"), and P(Squad ability to deal with injuries="High")	+2.81
P(EU competition="Champions League"), and P(EU readiness="No/Low")	+1.52
P(New Manager="No")	+0.92
P(Newly promoted="No (Low strength)")	+0.4
P(New Manager="Yes"), and P(Longevity of previous manager="Low")	-1.05
P(Net transfer spending...="Lower"), and P(Team wages...="High decrease")	-1.75
P(Newly promoted="No (High strength)")	-2.91
P(Newly promoted="No (Medium strength)")	-3.02
P(Changes in quality of players="Unstable")	-3.73
P(New Manager="Yes"), and P(Longevity of previous manager="High")	-4.64
P(Injury level="High"), and P(Squad ability to deal with injuries="Low")	-8.31
P(EU competition="Europa league"), and P(EU readiness="No/Low")	-8.47
P(EU competition="Both"), and P(EU readiness="No/Low")	-16.52

7. CONCLUSIONS

We have illustrated the reasoning towards the smart-data approach which is based on methods that support knowledge engineering and data engineering approaches to model development. The purpose of the knowledge engineering approach is to ensure that the model development is driven by what data are really required, rather than by what data are available, in order to achieve predictions and inferences of interest. The purpose of the data engineering approach is to ensure that data adhere to the causal (or otherwise) structure of the model, driven by real-world ‘facts’.

The paper demonstrates how we applied this process of reasoning to the problem of predicting changes in football team performances between seasons in a time-series approach. This enabled us to provide more accurate predictions of the evolution of football teams’ performance than is possible from purely data-driven models including standard non-linear regression models. Furthermore, the resulting model compares reasonably well against match prediction based models, including the bookmakers, which are based on far more data.

The implications of the paper are two-fold. First, with respect to the application domain, the current state-of-the-art is extended as follows:

1. This is the first study to present a model for time-series forecasting in terms of how the strength of football teams evolves over adjacent seasons, without the need to generate predictions for individual matches.
2. The results provide a novel and comprehensive attribution study of the factors most affecting performance (measured in terms of actual league points difference). For example, although unsurprisingly, the largest improvements in performance result from massive increases in spending on new players (an 8.49 points gain), an even greater decrease (up to 16.52 points) results from involvement in the European competitions (especially the Europa League) for teams that have previous little experience in such competitions.
3. Studies which assess the efficiency of the football gambling market may find the BN model helpful in the sense that it could help explain previously unexplained fluctuations in football odds.
4. Although the study had not intended to address it, the results also question (at least to some extent) the cause of the widely accepted favourite-longshot bias observed in bookies odds. This is because our results show that the favourite-longshot bias pattern holds for different types of models, not just for bookmakers’ odds, and arrives naturally irrespective of whether knowledge is incorporated with data.

Second, with respect to the general strategy for learning from data, we have demonstrated that seeking larger datasets is not always the solution in terms of increasing predictive accuracy. The model presented in this paper, for instance, is based on just 300 data instances generated over a period of 15 years, whereas some of the external models considered for comparison purposes, and which underperformed relative to the BN model, are based on up to 5,700 match instances over those 15 years. We have also shown that standard non-linear statistical regression based models, which are still the standard method used in many areas of data-driven prediction for critical real-

world risk assessment problems, such as in medical decision analysis (Kendrick, 2014), failed to achieve predictive accuracy similar to the smart-data BN, or any of the other match-by-match models.

Smart-data aims to improve the quality, as opposed to the quantity, of a dataset which also directly influences the quality of the model. With smart-data one has to have a clear understanding of the inferences of interest. Inferring knowledge from data imposes further challenges and requires skills that merge the quantitative as well as qualitative aspects of data.

The paper questions whether automated learning of the available data is capable of inferring real-world facts, such as those incorporated into the BN model presented in this paper. It may be the case that resulting inferences in many real-world problems will be limited in the absence of expert intervention for data engineering and structure modelling purposes. Future research will examine the capability of causal discovery algorithms in terms of realising various real-world facts, and the impact various data-engineering interventions may have on the results.

ACKNOWLEDGEMENTS

We acknowledge the financial support by the European Research Council (ERC) for funding this research project, ERC-2013-AdG339182-BAYES_KNOWLEDGE, and Agena Ltd for software support.

REFERENCES

- [1] Agena. (2016). Bayesian Network and Simulation Software for Risk Analysis and Decision Support. Retrieved online, November 11, 2016, from: <http://www.agenarisk.com>
- [2] Baio, G., & Blangiardo, M. (2010). Bayesian hierarchical model for the prediction of football results. *Journal of Applied Statistics*, 37:2, 253- 264.
- [3] Cain, M., Law, D., & Peel, D. (2000). The favourite-longshot bias and market efficiency in UK football betting. *Scottish Journal of Political Economy*, 47: 25–36.
- [4] Constantinou, A., Fenton, N., & Neil, M. (2012). pi-football: A Bayesian network model for forecasting Association Football match outcomes. *Knowledge-Based Systems*, 36: 322-339.
- [5] Constantinou, A., Fenton, N., & Neil, M. (2013a). Profiting from an inefficient association football gambling market: Prediction, Risk and Uncertainty using Bayesian Networks. *Knowledge-Based Systems*, 50: 60-86.
- [6] Constantinou, A. & Fenton, N. (2013b). Determining the level of ability of football teams by dynamic ratings based on the relative discrepancies in scores between adversaries. *Journal of Quantitative Analysis in Sports*, 9(1): 37-50.
- [7] Constantinou, A, & Fenton, N. (2013c). Profiting from arbitrage and odds biases of the European football gambling market. *The Journal of Gambling Business and Economics*, Vol. 7, 2: 41-70.
- [8] Constantinou, A., Fenton, N., & Pollock, L. J. H. (2014). Bayesian networks for unbiased assessment of referee bias in Association Football. *Journal of Psychology of Sport and Exercise*, Vol. 15, 5: 538-547.
- [9] Constantinou, A., Fenton, N., & Neil, M. (2016). Integrating Expert Knowledge with Data in Bayesian Networks: Preserving Data-Driven Expectations when the Expert Variables Remain Unobserved. *Expert Systems with Applications*, 56: 197-208.

- [10] Crowder, M., Dixon, M., Ledford, A., & Robinson, M. (2002). Dynamic modelling and prediction of English football league matches for betting. *The Statistician*, 51, 157–168.
- [11] Dixon, M. & Coles, S. (1997). Modelling association football scores and inefficiencies in the football betting market. *Applied Statistics*, 46(2), 265–280.
- [12] Dixon, M., & Pope, P. (2004). The value of statistical forecasts in the UK association football betting market. *International Journal of Forecasting*, 20, 697-711.
- [13] Elo, A. E. (1978). *The rating of chess players, past and present*. New York: Arco Publishing.
- [14] FIFA. (2016). FIFA/Coca-Cola World Ranking. Retrieved online, November 15, 2016, from: <http://www.fifa.com/fifa-world-ranking/procedure/men.html>
- [15] Football Data. (2016). Data Files: England. Retrieved online, January 6, 2017, from: <http://www.football-data.co.uk/englandm.php>
- [16] Forrest, D., & Simmons, R. (2001). *Globalisation and efficiency in the fixed-odds soccer betting market*. University of Salford, Salford: Mimeo.
- [17] Forrest, D., & Simmons, R. (2002). Outcome uncertainty and attendance demand in sport: the case of English soccer. *The Statistician*, 2, 241-291.
- [18] Forrest, D., Goddard, J., & Simmons, R. (2005). Odds-setters as forecasters: The case of English football. *International Journal of Forecasting*, 21, 551-564.
- [19] Forrest, D., & Simmons, R. (2008). Sentiment in the betting market on Spanish football. *Applied Economics*, 40, 119-126.
- [20] Friedman, N., Getoor, L., Koller, D., & Pfeffer, A. (1999). Learning Probabilistic Relational Models. In *Proceeding of the Sixteenth International Joint Conference of Artificial Intelligence (IJCAI-99)*, Stockholm, Sweden, August 1999.
- [21] Goddard, J., & Asimakopulos, I. (2004). Forecasting Football Results and the Efficiency of Fixed-odds Betting. *Journal of Forecasting*, 23, 51-66.
- [22] Graham, I., & Stott, H. (2008). Predicting bookmaker odds and efficiency for UK football. *Applied Economics*, 40, 99-109.
- [23] Hvattum, L. M., & Arntzen, H. (2010). Using ELO ratings for match result prediction in association football. *International Journal of Forecasting*, 26, 460-470.
- [24] IMF. (2009). World Economic Outlook: Crisis and Recovery. *International Monetary Fund*, April 2009.
- [25] Joseph, A., Fenton, N., & Neil, M. (2006). Predicting football results using Bayesian nets and other machine learning techniques. *Knowledge-Based Systems*, 7, 544-553.
- [26] Karlis, D., & Ntzoufras, I. (2003). Analysis of sports data by using bivariate Poisson models. *The Statistician*, 52: 3, 381-393.
- [27] Kendrick, M. (2014). *Doctoring Data: How to sort out medical advice from medical nonsense*. UK, Columbus Publishing.
- [28] Knorr-Held, L. (2000). Dynamic Rating of Sports Teams. *The Statistician*, 49(2), 261-276.
- [29] Kuypers, T. (2000). Information and efficiency: an empirical study of a fixed odds betting market. *Applied Economics*, 32, 1353-1363.
- [30] Leitner, C., Zeileis, A., & Hornik, K. (2010). Forecasting sports tournaments by ratings of (prob)abilities: A comparison for the EURO 2008. *International Journal of Forecasting*, 26, 471-481.

- [31] Maher, M. J. (1982). Modelling association football scores. *Statistica Neerlandica*, 36(3), 109–11.
- [32] McHale, I. G., Scarf, P. A., & Folker, D. E. (2012). On the development of a soccer player performance rating system for the English premier league. *Interfaces*, 42 (4), 339-351.
- [33] Min, B., Kim, J., Choe, C., Eom, H., & McKay, R. B. (2008). A compound framework for sports results prediction: A football case study. *Knowledge-Based Systems*, 21, 551-562.
- [34] MyFootballFacts (2016). Premier League Managers 1992-93 to 2015-16. Retrieved online, January 26, 2016 from: http://www.myfootballfacts.com/Premier_League_Managers.html
- [35] Neil, M., Tailor, M., & Marquez, D. (2007). Inference in hybrid Bayesian networks using dynamic discretization. *Statistics and Computing*, 17, 219-233.
- [36] Peel, D. A., & Thomas, D. A. (1988). Outcome uncertainty and the demand for football. *Scottish Journal of Political Economy*, 35, 242-249.
- [37] Peel, D. A., & Thomas, D. A. (1992). The demand for football: some evidence on outcome uncertainty. *Empirical Economics*, 4, 567-570.
- [38] Peel, D. A., & Thomas, D. A. (1997). Handicaps, outcome uncertainty and attendance demand. *Applied Economic Letters*, 4, 567-570.
- [39] Physio Room (2016). English Premier League Injury Table. Retrieved online, January 26, 2016, from: http://www.physioroom.com/news/english_premier_league/epl_injury_table.php
- [40] Pope, P., & Peel, D. (1989). Information, prices and efficiency in a fixed-odds betting market. *Economica*, 56, 323-341.
- [41] Rotshtein, A., Posner, M., & Rakytyanska, A. (2005). Football predictions based on a fuzzy model with genetic and neural tuning. *Cybernetics and Systems Analysis*, 41: 4, 619- 630.
- [42] Rue, H., & Salvesen, O. (2000). Prediction and retrospective analysis of soccer matches in a league. *The Statistician*, 3, 339-418.
- [43] Sauer, R. D. (1998). The Economics of Wagering Market. *Journal of Economic Literature*, 36, 2021-2064.
- [44] Telegraph (2016). Graphic: Premier League transfer and wage spending from the 2000/01 season to 2011/12. Retrieved online, January 26, 2016, from: <http://www.telegraph.co.uk/sport/football/competitions/premier-league/9514149/Graphic-Premier-League-transfer-and-wage-spending-from-the-20001-season-to-201112.html>
- [45] Total Sportek (2015). Premier League Player Salaries of 20 Clubs (Wage Bills 2015). Retrieved online, August 15, 2015, from: <http://www.totalsportek.com/money/english-premier-league-wage-bills-club-by-club/>
- [46] TSM Plug (2016). Premier League Player Salaries (Club by Club): All Premier League Clubs Wage Bill 2011 to 2014 Seasons. Retrieved online, January 26, 2016, from: <http://www.tsmplug.com/football/premier-league-player-salaries-club-by-club/>
- [47] Vecer, J., Kopriva, F., & 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
- [48] WhoScored (2016). Premier League Player Statistics. Retrieved online, January 26, 2016, from: <https://www.whoscored.com/Regions/252/Tournaments/2/Seasons/4311/Stages/9155/PlayerStatistics/England-Premier-League-2014-2015>
- [49] Wieland, V., & Wolters, M. (2012). Macroeconomic model comparisons and forecast competitions. VOX: CEPR's Policy Portal: Retrieved online, November 16, 2016, from: <http://voxeu.org/article/failed-forecasts-and-financial-crisis-how-resurrect-economic-modelling>

- [50] Wikipedia (2016a). 2000-01 FA Premier League. Retrieved online, January 26, 2016, from: https://en.wikipedia.org/wiki/2000%E2%80%9301_FA_Premier_League
- [51] Wikipedia (2016b). 2014-15 Arsenal F.C. season. Retrieved online, January 26, 2016, from: https://en.wikipedia.org/wiki/2014%E2%80%9315_Arsenal_F.C._season
- [52] XScores (2016). England – Premier League. Retrieved January 26, 2016, from: <http://www.xscores.com/soccer/tables>