Modeling and Simulation of Euro - 2016

Ninad Khargonkar[1], Saman Muthukumarana [a,2,3,*]

[a]Department of Statistics, University of Manitoba, Winnipeg, Manitoba R3T 2N2, Canada.

Abstract

In this paper we propose two simple simulation models for predicting the matches of Euro - 2016 tournament which is organized by the Union of European Football Associations. The two models differ in their assignment of winning probabilities for the matches. The simulation of the matches is used to analyze questions like the impact of the first group stage match on further qualification, the relative strength of the groups and analyzing any inherent bias in the knockout format of the tournament.

Keywords: Analytics, Simulation, Sports statistics, Bayesian network

1. Introduction

Association football (also known as 'soccer' or 'football') is the most popular sport in the world. One of the main reason for football's popularity is the uncertain nature of matches in tournaments like the World Cup and the Euro cup. Simulation of models can be helpful in making observations related to the results of matches and progress of teams in such tournaments. Researchers have proposed different types of simulation models which incorporate some form of a ranking system for the teams playing in the tournament (Koning et al. [1] and Leitner et al. [2]).

We will consider two models which differ in their assignment of probabilities for the outcome of a match (win / loss / draw). The outcome of a match can be classified into either a draw or not a draw. Therefore we will consider two probability distributions which take care of the either possibility for modeling the score of a match.

*Principal Corresponding author

Email addresses: ninad.khargonkar@gmail.com (Ninad Khargonkar), Saman.Muthukumarana@umanitoba.ca (Saman Muthukumarana)

[1]Ninad Khargonkar is a Research Intern at University of Manitoba under Mitacs Globalink program.

[2]Saman Muthukumarana is an Associate Professor in the Department of Statistics, University of Manitoba, Winnipeg, Manitoba R3T 2N2, Canada.

[3]Muthukumarana has been partially supported by grant from the Natural Sciences and Engineering Research Council of Canada. The authors thank Co-Editor-in-Chief and two anonymous reviewers whose comments helped to improve the manuscript.

The remainder of the paper is structured as follows: In Section 2 the models for predicting the outcome of matches are presented for the group stages and knockout stages of the tournament. They try to incorporate the form of a team in the tournament and propose a simple rule for updating the probabilities for future matches. In Section 3 the results and observations from the simulations of the tournament are presented. Each simulation of the tournament involved predicting the outcome and score for all 51 matches. The progress of each team was determined according to the tie-breaker rules of the tournament given by UEFA [3]. We conclude the report with a short discussion of drawbacks and potential improvements of the models in Section 4.

1.1. Description of the tournament

The UEFA (Union of European Football Associations) European Championship – 2016 or 'Euro – 2016' was different in size and format from previous Euro cups as it was expanded to a total of 24 teams. The teams were drawn into six groups as shown in Table 1. From these groups the 6 group winners, the 6 group runners – up and the 4 best third placed teams were selected for the knockout stages. In previous Euro cups which had a total of 16 teams, the top 2 teams from the 4 groups advanced directly to the quarter-finals of the tournament.

Table 1: **Groups of Euro – 2016**

Group A	**Group B**	**Group C**	**Group D**	**Group E**	**Group F**
France	England	Germany	Spain	Belgium	Portugal
Romania	Russia	Ukraine	Czech Rep.	Italy	Iceland
Albania	Wales	Poland	Turkey	Rep. of Ireland	Austria
Switzerland	Slovakia	Northern Ireland	Croatia	Sweden	Hungary

2. Model Development

We will consider two different models for the assignment of probabilities. The first model uses the data from the qualifying rounds of Euro – 2016 along with the Elo rating for each team at the start of the tournament. The Elo rating system has been found to be a good ranking system for teams (Lasek et al. [4]). It places more importance on the recent performances of each team. As France qualified automatically as a consequence of hosting the tournament, the data for the first model was taken from the international friendlies played by France during the duration of the qualifiers.

The second model uses the data from the previous matches played between all the teams. Hence all the 276 $\{= \binom{24}{2}\}$ possible combinations between the teams were recorded. Each record summarized the total number of games played between the teams and the number of wins, losses and draws between them. In case of a combination where the two teams had not played any match before, the missing data was taken from the qualifying stages of those teams (there were only 5 such pairs).

2.1. Group Stages

Bayesian networks can provide an effective framework to model a match or the group stages of a tournament and the subsequent progress to the knockout stages as seen in the papers by Constantinou et al. [5] and Joseph et al. [6]. The results of the matches in the group stages can influence the future performances.

For a group consisting of 4 teams, the six $\{\binom{4}{2}\}$ matches for it are played over the course of 3 different match-days. On a particular match-day two games are played between the two pairs of teams from the group. The influence network between the matches for the group stage is shown in Fig. 1.

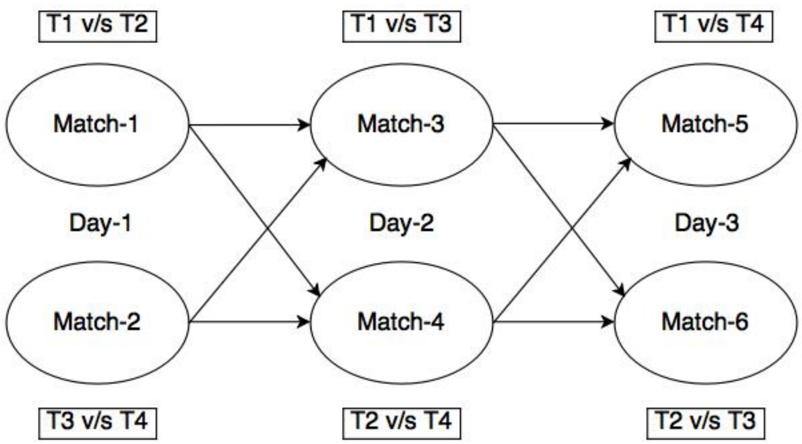

Figure 1: **The influence network for group stage**

Let X_i denote the outcome of the i^{th} match of the group (between teams T_{i1} and T_{i2}). For example the variable X_3 will denote the outcome of the third match of the group (T_1 v/s T_3). The assignment for X_i based on the outcome of the match will be as follows:

$$X_i = \begin{cases} 3 & \text{if } T_{i1} \text{ wins} \\ 0 & \text{if } T_{i1} \text{ loses} \\ 1 & \text{if draw occurs} \end{cases}$$

This assignment is helpful in updating the total number of points obtained by a team which is needed for determining the position of the team in the group.

For the first model, if g, w, l, d represent the number of games played, wins, losses and draws in the qualifiers then the following probabilities are assigned. Here $scaled(elo.T_1)$ denotes the scaled Elo rating for team T_1. $\{scaled(elo.T_1) = \frac{elo.T_1}{elo.T_1+elo.T_2}\}$

$P($ T_1 wins $) = \{(w_1 + l_2)/(g_1 + g_2)\} * scaled(elo.T_1)$
$P($ T_2 wins $) = \{(w_2 + l_1)/(g_1 + g_2)\} * scaled(elo.T_2)$
$P($ Draw $) = (d_1 + d_2)/(g_1 + g_2)$

In case of the second model, the probabilities can be found out from the historical record table by looking at the total number of wins, losses and draws for each of the 276 possible team combinations. It will be similar to the first model but will give weight to the historical performances of the teams.

The probability for the outcomes of the first two matches i.e X_1 and X_2 (which are played on the first match day), is directly taken from the models and is not conditional upon anything. For the outcomes of the matches in second and third match day, a conditional probability distribution based on the two matches from previous match day is constructed. Therefore, X_3 and X_4 are conditional upon values of X_1 & X_2 and X_5 & X_6 are conditional upon values of X_3 and X_4.
For calculating the conditional distribution for a match between T_{i1} and T_{i2}, the following cases from previous match-day are considered:

1. Both of them had same result on the previous match-day.
2. T_{i1} performed better than T_{i2}.
3. T_{i2} performed better than T_{i1}.

In the first case, the probability for a draw is increased as both the teams had similar results. In second and third case, the winning probability for the team having a better result on previous match day is increased and the win probability for the other team is accordingly reduced.

For example if p_1, p_2, p_3 represent the unconditional probabilities for the win, loss and draw for a certain team and i^{th} match i.e $P(X_i = 3) = p_1$, $P(X_i = 0) = p_2$ and $P(X_i = 1) = p_3$ (taken from the record table of either model) then we obtain the conditional probabilities from the following update rules:

if Case(1) **then**
    $p3 \leftarrow p3 + \frac{(p1+p2)}{4}$
    $p1 \leftarrow 0.75 * p1$
    $p2 \leftarrow 0.75 * p2$
end if
if Case (2) **then**
    $p1 \leftarrow p1 + \frac{p2}{4}$

$p2 \leftarrow 0.75 * p2$
end if
if Case (3) **then**
$p2 \leftarrow p2 + \frac{p1}{4}$
$p1 \leftarrow 0.75 * p1$
end if

2.1.1. Score of a match

The scores from all the matches played at the previous Euro cups were recorded for building the probability distribution for each type of score (draw / non draw). Four types of scores were considered when the result was a draw as shown in Table 2 and nine types of scores were considered when the result was not a draw as shown in Table 3.

Table 2: **Match was a draw**

Score Type	% chance
0-0	40
1-1	48.33
2-2	8.33
3-3	3.33

Table 3: **Match was not a draw**

Score Type	% chance
1-0	24.85
2-1	24.86
3-2	9.46
2-0	19.52
3-1	6.51
4-2	1.77
3-0	9.46
4-1	2.36
4-1	1.18

2.2. Knockout stages

As a consequence of the tournament structure, the position of the team in its group (after the group stage was over) had a direct influence on the result of its knockout match. The team having a higher position in its group would possibly face a weaker opposition as the opposition would have a lower position in its group. This resulted in a relatively simple influence diagram as shown in Fig. 2.

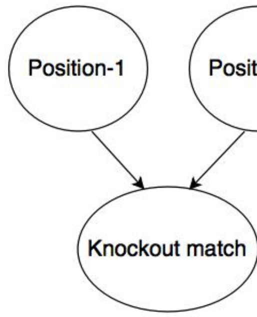

Figure 2: **The influence diagram for knockout rounds**

The importance given to the difference in position of the teams was less that in the group stage because some groups may be very competitive ('Group of Death') and other groups may be relatively one-sided.
If p_1, p_2, p_3 represent the unconditional probabilities for team1 - win, team2 - win and a draw respectively then the conditional probabilities would be updated by the following rules:

if Case(1): (Same position for both) **then**
  No change
end if
if Case (2): (Position(T_1) > Position(T_2)) **then**
  $p1 \leftarrow p1 + \frac{p2}{5}$
  $p2 \leftarrow 0.8 * p2$
end if
if Case (3): (Position(T_2) > Position(T_1)) **then**
  $p2 \leftarrow p2 + \frac{p1}{5}$
  $p1 \leftarrow 0.8 * p1$
end if

2.3. Penalties in knockout stage

If a knockout stage match resulted in a draw, a penalty shootout was simulated between the teams. Instead of giving a 50 – 50 chance to the teams, a different approach was used on the basis of the paper by Apesteguia and Palacios-Huerta [7]. The authors of the paper concluded that the team taking into penalties first had a 60.5% of winning the shootout. This is also mentioned in the popular book 'Soccernomics' (Kuper [8]). A fair-coin toss decided the team taking the penalty kicks first. Then the probabilities for the penalty shootout were defined as follows :

P (T_1 wins | T_1 is first) = 0.605
P (T_1 wins | T_1 is second) = 0.395

The four best-placed teams are:	WA plays	WB plays	WC plays	WD plays
A B C D	3C	3D	3A	3B
A B C E	3C	3A	3B	3E
A B C F	3C	3A	3B	3F
A B D E	3D	3A	3B	3E
A B D F	3D	3A	3B	3F
A B E F	3E	3A	3B	3F
A C D E	3C	3D	3A	3E
A C D F	3C	3D	3A	3F
A C E F	3C	3A	3F	3E
A D E F	3D	3A	3F	3E
B C D E	3C	3D	3B	3E
B C D F	3C	3D	3B	3F
B C E F	3E	3C	3B	3F
B D E F	3E	3D	3B	3F
C D E F	3C	3D	3F	3E

Figure 3: **The possible parings for third placed teams in Round of 16**
(Image was taken from the regulations document from Euro 2016 official website of UEFA [3])

3. Results from simulation

The models discussed in Section 2 were implemented in the R programming environment to perform the simulations. Ten thousand simulations were performed for each type of model to obtain the results. The final position of each team in its group was determined after simulating the results and scores of all the thirty-six matches in the group stage. The subsequent qualification of the teams to the knockout stages was determined according to the rules of the tournament. The observations and results are shown in the following subsections.

3.1. Bias in Round of 16 pairings

Teams from some groups are at a disadvantage when the matches in the Round of 16 stage are decided (Fig. 3). The four group-winners from groups A, B, C, D are matched with the third-placed teams from other groups while winners from Groups E and F face second-paced teams from other groups. In addition the Group E runner up has to face a group winner.

Due to such a bias it is also possible that winners from Groups A and D do not face another group winner until the semi-final thereby making their path to the final relatively easier. This way of pairing is very unbalanced and a much better solution would be to rank winners and runner ups in the same way as it was done for third placed teams and then pair-up the teams based on the rankings.

Both the models were able to point out the inconsistencies described above. As we can see in Fig. 4, groups E and F clearly fall behind others on comparing the relative proportion of quarter-finalists from each group. In the second model group F lags behind even more when compared with the first model.
The first model is better at pointing out the bias and confirming the disadvantage towards group E (both its runner and winner are at a loss).

On observing the progress of teams in the first model after the quarterfinals and into the semi-finals and the final as shown in Fig. 5, Groups B and D perform better than every other group indicating the possibility of an easier path to final.

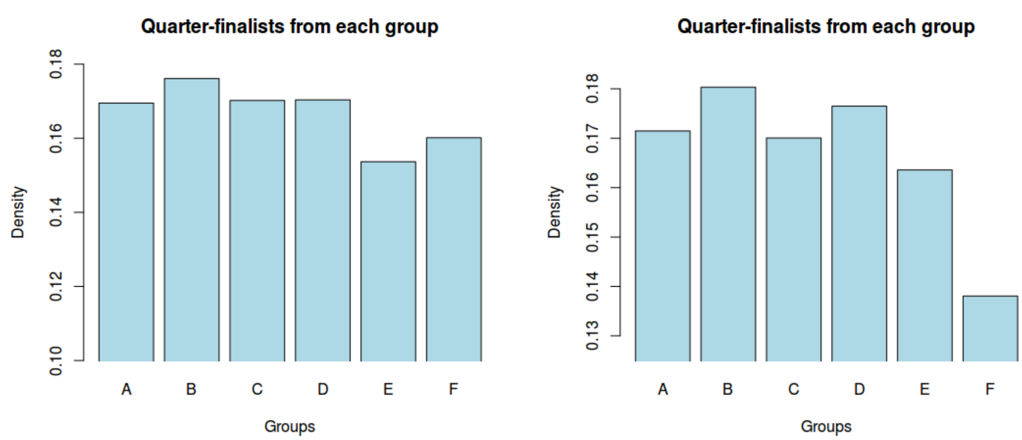

Figure 4: **Proportion of Quarter finalists from the first and second model**

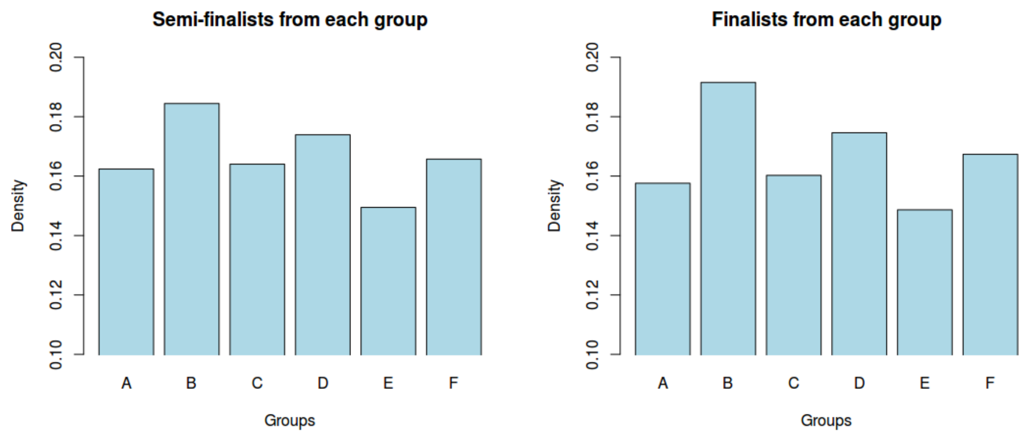

Figure 5: **Semi - finalists and Finalists from the second model**

3.2. Probability for Qualification

We also take a look at the probability for each team to qualify for the knockout stages. This gives us an idea about the relative strength of the groups. For each model, the five teams with the highest probabilities for qualification are listed in Table 4 and Table 5. The plot for the probabilities of all the teams is given in Fig. 6.

Table 4: **Model – 1**

Team	% chance
England	85.48
Italy	80.33
Austria	79.80
Spain	79.59
Belgium	78.70

The second model gives higher values for the qualification probabilities and it also has the traditionally strong teams at the top. As the first model incorporates the data from the qualifiers of the tournament, it may have a bias towards teams that performed well during tournament qualification. For example, Belgium shows up in top five with first model but lags far behind in the second model. This may be due to its current team performing very well in the qualifiers when compared to its previous teams which were not that significant.

Table 5: **Model – 2**

Team	% chance
Portugal	97.56
Germany	94.28
England	94.16
France	88.25
Switzerland	88.05

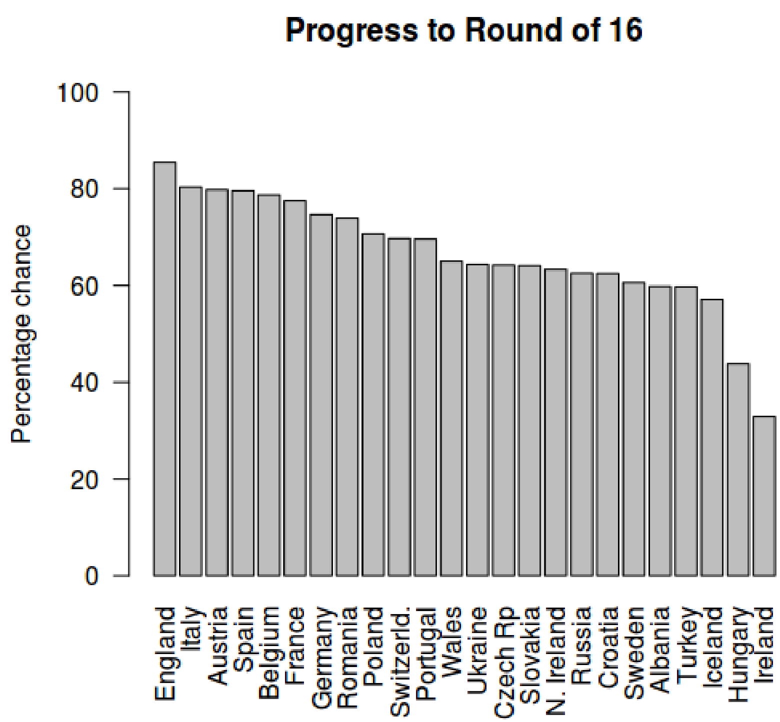

Figure 6: **Model – 1**

3.3. Impact of the first game

A lot importance and scrutiny is placed on a team's performance in its first match of the the tournament. It is variable which can affect the team's morale and performance on other match-days. We take a look at how winning or losing the first match can have an impact on subsequent qualification in Fig. 7 and 8. Some teams show lower values for qualification probabilities on winning first match when compared to losing the first match. This may indicate that they had very strong opponents from the group on the first day itself.
But if we take into account the overall picture, a win in the first match gives a much better chance for qualification to the knockout rounds.

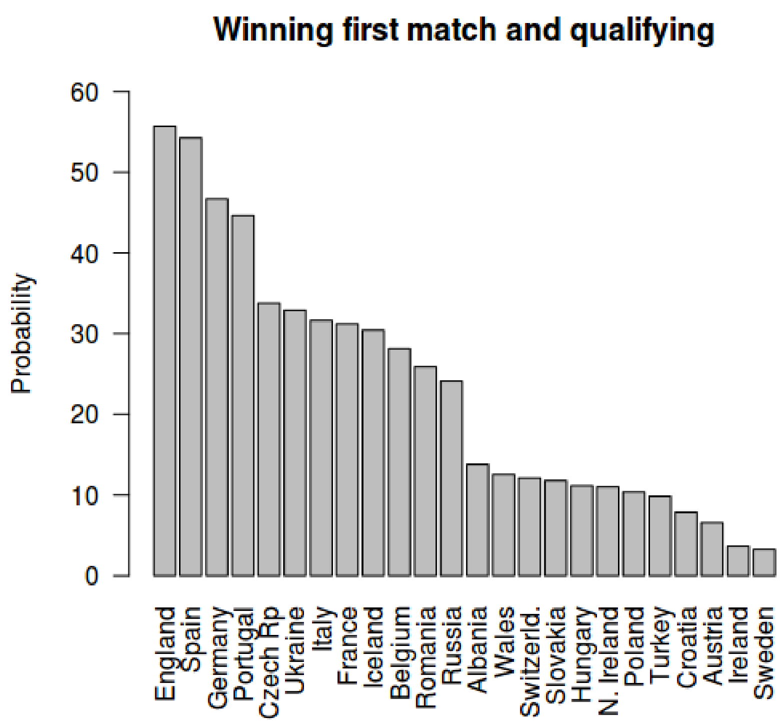

Figure 7: **Model – 1**

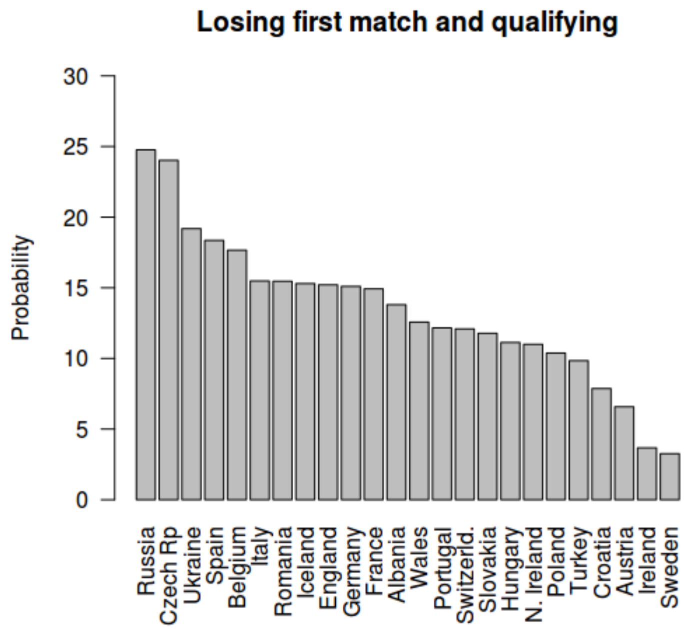

Figure 8: **Model – 1**

4. Discussion and Conclusions

In this paper, two simple and easy to implement models for simulating matches of Euro – 2016 are proposed. The models incorporate the recent performances of the teams as well as their historical record. For the progress of the teams, various rules based on previous results in the group stage are introduced for updating the probability of outcome of the match. From the results of the simulation, certain groups are shown to be at a disadvantage in the knockout stages due to the nature of the tournament format.

Although the models show the disadvantages faced by some teams in knockout stages, they are not good enough for the prediction of the knockout rounds due to their simplicity and the randomness introduced by the penalty shootout.

Both the models tend to give very uniform probabilities for success of teams in the knockout rounds which is highly unlikely. Each model is affected by the simplicity of assumptions which results in no team being the outright favorite to win the tournament.

In the case of simulation of a single game, the assignment of the probability can be improved by using complex Bayesian networks which taken into account a lot of team specific variables like injuries and player quality. Other game related statistics shown in the paper by Lago-Peñas et al. [9] could also be helpful in preparing a complex model for the simulation.

References

References

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

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

[3] UEFA, Euro – 2016 competition rules and format, ???? `http://www.uefa.com/uefaeuro/about-euro/format/index.html`.

[4] J. Lasek, Z. Szlávik, S. Bhulai, The predictive power of ranking systems in association football, International Journal of Applied Pattern Recognition 1 (2013) 27–46.

[5] A. C. Constantinou, N. E. Fenton, M. Neil, pi-football: A bayesian network model for forecasting association football match outcomes, Knowledge-Based Systems 36 (2012) 322–339.

[6] A. Joseph, N. E. Fenton, M. Neil, Predicting football results using bayesian nets and other machine learning techniques, Knowledge-Based Systems 19 (2006) 544–553.

[7] J. Apesteguia, I. Palacios-Huerta, Psychological pressure in competitive environments: Evidence from a randomized natural experiment, The American Economic Review 100 (2010) 2548–2564.

[8] S. Kuper, Soccernomics: Why England Loses, Why Spain, Germany, and Brazil Win, and Why the US, Japan, AustraliaÑand Even IraqÑAre Destined to Become the Kings of the World's Most Popular Sport, Nation Books, 2014.

[9] C. Lago-Peñas, J. Lago-Ballesteros, A. Dellal, M. Gómez, Game-related statistics that discriminated winning, drawing and losing teams from the spanish soccer league, Journal of Sports Science and Medicine 9 (2010) 288–293.