Communications for Statistical Applications and Methods
2016, Vol. 23, No. 4, 297–319
http://dx.doi.org/10.5351/CSAM.2016.23.4.297
Print ISSN 2287-7843 / Online ISSN 2383-4757

Predicting football scores via Poisson regression model: applications to the National Football League

Erlandson F. Saraiva[a], Adriano K. Suzuki[1,b], Ciro A. O. Filho[b], Francisco Louzada[b]

[a]Institute of Mathematics, Federal University of Mato Grosso do Sul, Brazil;
[b]Department of Applied Mathematics and Statistics, University of São Paulo, Brazil

Abstract

Football match predictions are of great interest to fans and sports press. In the last few years it has been the focus of several studies. In this paper, we propose the Poisson regression model in order to football match outcomes. We applied the proposed methodology to two national competitions: the 2012–2013 English Premier League and the 2015 Brazilian Football League. The number of goals scored by each team in a match is assumed to follow Poisson distribution, whose average reflects the strength of the attack, defense and the home team advantage. Inferences about all unknown quantities involved are made using a Bayesian approach. We calculate the probabilities of win, draw and loss for each match using a simulation procedure. Besides, also using simulation, the probability of a team qualifying for continental tournaments, being crowned champion or relegated to the second division is obtained.

Keywords: prediction, football, attack and defense effect, Poisson regression, Bayesian inference, MCMC, simulation, de Finetti measure

1. Introduction

Football, originally practiced in England, is one of the most popular collective sports worldwide. A particular characteristic of this sport is that the best team it is not always the winner of a match or a tournament, which causes a climate of expectation among players and fans.

In the last few years, some studies have addressed the prediction of outcomes for matches of the World Cup, such as, Dyte and Clarke (2000), Volf (2009), and Suzuki *et al.* (2010). Dyte and Clarke (2000) proposed a Poisson regression model considering control variables, which consist of the rating for each team and the match venue given by the Federation Internationale of Football Association (FIFA). The authors used their results and other results about the quality of forecasts to simulate the 1998 FIFA World Cup. Volf (2009) consider a counting processes approach, in order to model a match score as two interacting time-dependent random point process. The interaction between teams are modeled via a semi-parametric multiplicative regression model of intensity. The authors applied this model to the analysis of the performance of the eight teams that reached the quarter-finals of the 2006 FIFA World Cup.

Suzuki *et al.* (2010), proposed a Bayesian approach to predict of the outcomes of matches using specialists' opinions and FIFA rankings to build a Power *prior*. Using simulations, the authors calculate the probabilities of wins, draws, losses and odds of the teams being ranked in the group stage

[1] Corresponding author: Department of Applied Mathematics and Statistics, University of São Paulo, Avenida Trabalhador São-carlense, 400 - Centro CEP: 13566-59, São Carlos-SP, Brazil. E-mail: suzuki@icmc.usp.br

Published 31 July 2016 / journal homepage: http://csam.or.kr
© 2016 The Korean Statistical Society, and Korean International Statistical Society. All rights reserved.

are obtained. Bastos and da Rosa (2013) developed a Bayesian methodology for the Poisson-gamma model in which the *priors* are chosen considering historical and recent information. The authors calculate the probabilities of win, draw and loss for the 2010 FIFA World Cup games.

Several articles also focused on the prediction of outcomes in national leagues. Among them, Keller (1994) considered the Poisson distribution for the number of goals scored by England, Ireland, Scotland and Wales in the British International Championship (1883–1980). Lee (1997) developed a generalized linear model with application to final rank analysis. Brillinger (2008) modeled the probabilities of win, tie and loss through an ordinal-value model and applied the model to the Brazilian Series A championship. Karlis and Ntzoufras (2009) applied the Skellam's distribution to model the difference of goals between home and away teams. The authors illustrated the model using the 2006–2007 English Premier League. Koopman and Lit (2015) developed a statistical model to predict the games of the 2010–2011 and 2011–2012 English Premier Leagues, assuming a bivariate Poisson distribution with coefficients that stochastically changed intensity over time.

An issue about papers cited above is that none consider the home team factor to calculate the probabilities of interest. For Maher (1982) it is important to add a constant factor to all teams when they play at home. Following this approach, Dixon and Coles (1997) presented a study considering 6,000 matches of English teams in the 1993–1995 period. Results showed that 46% of the matches was won by the home team, 27% were draws and in 27% the home team lost. In a similar study, Knorr-Held (2000) provided data of the 1996–1997 season of the German Bundesliga. Results showed that in 51% of the matches the win was of the home team and in only 26% the home team lost. Considering the season 2011–2012 of the English Premier League, 47% of the matches ended with win of the home team and only 27% ended with defeat of the home team. These results show us that, for some reason, there seems to be an inherent advantage for the team if it is playing at home. In this way, the effect of playing at home can be introduced in the model in order to predict the probabilities of win, draw and lose.

In this paper, we model the number of goal scored by each team in a match by a Poisson distribution, whose average reflects the strength of the attack and defense of the team and effect of being playing at home. Inferences about all the unknowns quantities involved are made using a Bayesian approach. We illustrate the performance of the proposed method considering the outcomes of the 2012–2013 season of the English Premier League (EPL) and the outcomes of the 2015 Brazilian Football League (BFL).

Using a simulation study, we calculate the probabilities of win, draw and lose for each team in each round of the EPL and BFL. We also present the probability of a team qualifying for the continental tournaments, being crowned champion or relegated to the second division. All computer implementations were performed using OpenBUGS (Spiegelhalter *et al.*, 2003) and R systems (R Development Core Team, 2012) in the R2WinBUGS package (Gelman *et al.*, 2006).

The remainder of the paper is organized as follows. In Section 2, we present the Poisson regression model and expressions used to calculate the probabilities of win, draw and defeat for a football game. Sections 3 and 4 report results obtained by applying the proposed model for matches of the EPL and BFL, respectively. Section 5 concludes with some general remarks.

2. Model

Consider a football championship with $n + 1$ teams, in which, each team plays $2n$ times, being n times at home stadium and n times at away stadium. The number of games of the championship is $N = n(n + 1)$. The N games are played in two phases, each phase with $N/2$ games. If in the first phase, a game between teams t and s occurs at home stadium of t team, then in the phase two the game occurs at home stadium of s team. By each result, victory, draw and defeat each team gets 3, 1 and 0 points, respectively. After the N games, the team with the highest score is declared champion. The M teams with smallest scores are relegated to the second division. In the EPL, $M = 3$, and in BFL, $M = 4$.

For a game j between teams t and s, let X_{tj} and X_{sj} be random variables denoting the number of goals of the home team and away team, respectively, for $j = 1, \ldots, n$. Assume that,

$$X_{tj} \sim \text{Poisson}(\lambda_{tj}) \quad \text{and} \quad X_{sj} \sim \text{Poisson}(\lambda_{sj}). \tag{2.1}$$

In order to link the number expected of goals of teams t and s with their strength of attack (a), strength of defense (d) of the opposing team and the effect of being playing at home (h), we consider

$$\lambda_{kj} = e^{U_{kj}\boldsymbol{\beta}_k}, \tag{2.2}$$

for $j = 1, \ldots, 2n$, where $k = t, s$, $U_{kj} = (1, 1, 1, 1)$ if the game is at home stadium of k team and $U_{kj} = (1, 1, 1, 0)$ otherwise, and $\boldsymbol{\beta}_k = (\beta_{k0}, \beta_{ka}, \beta_{k^c d}, \beta_{kh})'$ is the vector of parameters of the k team, where k^c represents the opposing team. For example, if $k = t$, then $k^c = s$. The parameter β_{ka} measures the attack strength of the k team and parameter $\beta_{k^c d}$ measures the defense strength of the opposing team k^c. The parameter β_{kh} gives the advantage of playing at home, which we assume as being equals for every team of the championship. Note that, in this formulation a team with a good defense will have a negative defense effect because this will decrease the expected number of goals of the opposing team. In the other hand, a team with positive defense effect increases the expected number of goals of the opponent.

Suppose that a game j is played in $(r + 1)^{th}$ round of the championship, $1 \leq r \leq N$. Let n_r be the number of games played by teams t and s before of the $(r + 1)^{th}$ round. Consider $\mathbf{x}_k = (x_{k1}, \ldots, x_{kn_r})$ be the number of goals scored by k team in the n_r games, in which, x_{km} is number of goals scored by the k team in the m^{th} game, for $k = t, s$ and $m = 1, \ldots, n_r$. Thus, the log-likelihood function for $(\boldsymbol{\beta}_t, \boldsymbol{\beta}_s)$ is given by

$$l(\boldsymbol{\beta}_t, \boldsymbol{\beta}_s; \mathbf{x}_t, \mathbf{x}_s) = \sum_{k \in \{t,s\}} \sum_{m=1}^{n_r} \left(-e^{U_{km}\boldsymbol{\beta}_k} + x_{km} U_{km} \boldsymbol{\beta}_k - \log(x_{km}!) \right). \tag{2.3}$$

Some constraints must be imposed on team-specific parameters to avoid nonidentifiability. Following Karlis and Ntzoufras (2003) and Baio and Blangiardo (2010), we use a sum-to-zero constraint, i.e.,

$$\sum_{t=1}^{n+1} \beta_{ta} = 0, \quad \sum_{t=1}^{n+1} \beta_{td} = 0, \quad \text{and} \quad \sum_{t=1}^{n+1} \beta_{th} = 0, \tag{2.4}$$

i.e., the sum of the strength of the attack, defense and home effect of all $(n + 1)$ teams is equal to zero.

In order to develop the Bayesian approach we need to specify the prior distributions for parameters $\boldsymbol{\beta}_k$, $k = t, s$. We assume that priors are *a priori* independent, i.e., $\pi(\boldsymbol{\beta}_t, \boldsymbol{\beta}_s) = \pi(\boldsymbol{\beta}_t)\pi(\boldsymbol{\beta}_s)$, in which, $\pi(\boldsymbol{\beta}_k) = \pi(\beta_{k0})\pi(\beta_{ka})\pi(\beta_{kd})\pi(\beta_{kh})$, for $k = t, s$. So, we consider the following prior distributions: $\beta_{k0} \sim \mathcal{N}(0, 10^{-4})$, $\beta_{ka} \sim \mathcal{N}(0, 10^{-3})$, $\beta_{k^c d} \sim \mathcal{N}(0, 10^{-3})$ and $\beta_{kh} \sim \mathcal{N}(0, 10^{-3})$, for $k = t, s$, where $\mathcal{N}(0, b)$ denotes the normal distribution with mean 0 and precision b.

Joint posterior distributions for parameters do not have closed form; therefore, we estimate parameters $\boldsymbol{\beta}_t$ and $\boldsymbol{\beta}_s$ using MCMC. In Appendix A of the Supplementary Material (SM) we provide some details of the estimation procedure using MCMC. All computer implementations were performed using OpenBUGS and R systems in the R2WinBUGS package. Estimates $\tilde{\boldsymbol{\beta}}_t$ and $\tilde{\boldsymbol{\beta}}_s$ are given by the average of the generated MCMC sample. Given $\tilde{\boldsymbol{\beta}}_t$ and $\tilde{\boldsymbol{\beta}}_s$, we use these values to calculate the probability of a win, draw and defeat of each team in the next round.

2.1. Predictions

Consider that a game j between teams t and s will occurs at home stadium of team t in $(r+1)^{th}$ round of the championship. Denote the probability of win, draw and defeat (loss) of team t by P_w, P_d and P_l, respectively. These probabilities are given by

$$P_w = P\left(X_{tj} > X_{sj} | \tilde{\boldsymbol{\beta}}_t, \tilde{\boldsymbol{\beta}}_s\right) = \sum_{g=1}^{\infty} \sum_{u=0}^{g-1} P\left(X_{tj} = g | \tilde{\boldsymbol{\beta}}_t\right) P\left(X_{sj} = u | \tilde{\boldsymbol{\beta}}_s\right), \tag{2.5}$$

$$P_d = P\left(X_{tj} = X_{sj} | \tilde{\boldsymbol{\beta}}_t, \tilde{\boldsymbol{\beta}}_s\right) = \sum_{g=0}^{\infty} P\left(X_{tj} = g | \tilde{\boldsymbol{\beta}}_t\right) P\left(X_{sj} = g | \tilde{\boldsymbol{\beta}}_s\right), \tag{2.6}$$

$$P_l = P\left(X_{tj} < X_{sj} | \tilde{\boldsymbol{\beta}}_t, \tilde{\boldsymbol{\beta}}_s\right) = \sum_{u=1}^{\infty} \sum_{g=0}^{u-1} P\left(X_{tj} = g | \tilde{\boldsymbol{\beta}}_t\right) P\left(X_{sj} = u | \tilde{\boldsymbol{\beta}}_s\right). \tag{2.7}$$

Similarly to Bastos and da Rosa (2013) and Suzuki *et al.* (2010) we calculate the de Finetti distance in order to measure the goodness of a prediction. This distance is given by the Euclidean distance between the point corresponding to the real outcome and the corresponding to the prediction. For this, is assumed that the set of all possible forecasts is given by the simplex set $\mathcal{S} = \{(P_w, P_d, P_l) \in \mathbb{R}^3 : P_w + P_d + P_l = 1, P_w \geq 0, P_d \geq 0, P_l \geq 0\}$ and that the possible real outcome, win, draw and defeat are represented by the points $(1,0,0)$, $(0,1,0)$ and $(0,0,1)$, respectively.

The de Finetti measure (df) is defined as:

$$\mathrm{df} = (P_w - b_1)^2 + (P_d - b_2)^2 + (P_l - b_3)^2,$$

where $(b_1, b_2, b_3) \in \{(1,0,0), (0,1,0), (0,0,1)\}$. For example, if the prediction for the game between teams t and s is $(0.2, 0.65, 0.15)$ and the real outcome is $(0,1,0)$, i.e., a draw, then the de Finetti distance is $df = (0.2-0)^2 + (0.65-1)^2 + (0.15-0)^2 = 0.185$.

For the equiprobable case, $P_w = P_d = P_l = 1/3$, with win of the home team, $(1,0,0)$, the de Finetti measure is given by $df = (1/3-1)^2 + (1/3-0)^2 + (1/3-0)^2 = 2/3$. This value is accepted as a threshold value in order to classify the predictions as acceptable or not, see for example Suzuki *et al.* (2010). If $df < 2/3$, the predictions are considered acceptable; otherwise, $df > 2/3$, the predictions are considered poor.

Using a simulation procedure, we also calculate the probability of each team to be the champion. In order to calculate these probabilities we assume that the first phase of the champion is ended, i.e., $n(n+1)/2$ games were played. Let T_t be the number of points of the team t until the last game of the first phase. The simulation procedure is given by the following steps:

(i) For the j^{th} game of the second phase of the championship, $j = (n(n+1)/2) + 1, \ldots, N$, do as follows:

   (a) Get the estimates $(\tilde{\boldsymbol{\beta}}_t, \tilde{\boldsymbol{\beta}}_s)$ from Bayesian approach, where t and s represent the teams t and s;

(b) Given $\tilde{\boldsymbol{\beta}}_k$, generate the number of goals scored by k team, X_{kj}, from a Poisson distribution with parameter $\lambda_{kj} = e^{U_{kj}\tilde{\boldsymbol{\beta}}_k}$, for $k = t, s$;

(c) If $X_{tj} > X_{sj}$, do $T_t = T_t + 3$ and $T_s = T_s$; if $X_{tj} = X_{sj}$, do $T_t = n_T + 1$ and $T_s = T_s + 1$; if $X_{tj} < X_{sj}$, do $T_t = T_t$ and $T_s = T_s + 3$;

(d) Ended the second phase, the A team is declared champion if $T_A = \max_{1 \le t \le n} T_t$.

(ii) Repeat the step (i), r times, for $r = 1, \dots, R$.

We consider $R = 1{,}000$. The probability of the team A be the champion is estimated by the proportion of times that A team is declared champion among the R simulated cases, i.e.,

$$P_{champ}(A) = \frac{N_A^{champ}}{R},$$

where N_A^{champ} is the number of times that team A is the champion among the R simulated cases.

Similarly, the probability of team A be relegated to the second division is given by

$$P_{releg}(A) = \frac{N_A^{releg}}{R},$$

where N_A^{releg} is the number of times that team A finished as one of the M teams with smaller number of points among the R simulated cases. For EPL, $M = 3$, and for BFL, $M = 4$.

3. Application 1

In this section, we apply the proposed method to the 2012–2013 season of the EPL. EPL is composite by $n + 1 = 20$ teams. The number of games of the EPL is 380, being 190 by phase.

Table 1 shows the number of games and the number of goals scored by each team at home and away in each phase of the EPL. Table 1 is ordered according to team with highest to smallest number of goals scored (last column). Manchester United has the highest number of goals scored, 86, being 45 scored at home and 41 away. Manchester United is also the team that has the highest number of goals scored away. In its home stadium, Manchester United, only scored less goals than Arsenal. Queens Park Rangers has the smallest number of goals scored, 30. This team, also has the smallest number of goals scored at home. West Ham United has the smallest number of goals scored at away.

Table 2, shows the number of games ended with number of goals (x_t, x_s), where x_t and x_s are the number of goals scored by home and away team, respectively. The sum of numbers of each column of this Table 2, give the number of games in which the home team scored x_t goals. For instance, the sum of the second column is 120; meaning, that in 120 games the home team scored 1 goal. Analogously, the sum of numbers of each line of Table 2, give the number of games in which the away team scored x_s goals. For instance, the sum of the fifth row is 15; meaning, that in 15 games the away team scored four or more goals. The main diagonal give the amount of ties. The total of ties is 109 (28.68%). The upper and lower diagonal give the number of games with win of the home and away team, respectively. Adding values of upper diagonal, we get the number of games won by home team, 166 (43.68%); while the sum of the values of the lower diagonal give the number of games won by the away team, 105 (27.64%).

Table 1: Number of games and goals scored by each team at home and away in each phase of the EPL

Team	1st phase				2nd phase				Number of goals		
	Games at home	Goals	Games away	Goals	Games at home	Goals	Games away	Goals	At home	Away	Total
Manchester United	09	26	10	22	10	19	09	19	45	41	86
Chelsea	10	24	09	15	09	17	10	19	41	34	75
Arsenal	09	23	10	14	10	24	09	11	47	25	72
Liverpool	10	14	09	14	09	19	10	24	33	38	71
Manchester City	10	22	09	12	09	19	10	13	41	25	66
Tottenham Hotspur	10	14	09	20	09	15	10	17	29	37	66
Everton	09	16	10	16	10	17	09	08	33	22	55
West Bromwich Albion	10	16	09	12	09	16	10	09	32	21	53
Fulham	09	16	10	13	10	12	09	09	28	22	50
Southampton	10	14	09	11	09	12	10	12	26	23	49
Aston Villa	09	08	10	07	10	15	09	17	23	24	47
Swansea City	10	17	09	10	09	11	10	09	28	19	47
Wigan	10	14	09	06	09	12	10	15	26	21	47
Newcastle United	10	12	09	11	09	12	10	10	24	21	45
West Ham United	10	17	09	06	09	17	10	05	34	11	45
Reading	09	14	10	07	10	09	09	10	23	20	43
Norwich City	09	10	10	10	10	15	09	06	25	16	41
Sunderland	09	10	10	10	10	10	09	11	20	21	41
Stoke City	09	11	10	07	10	10	09	06	21	13	34
Queens Park Rangers	09	08	10	08	10	05	09	09	13	17	30

Table 2: Number of games ended with number of goals (x_t, x_s), for $x_t, x_s \in \{0, 1, 2, 3, 4+\}$

Away team	Home team					Total
	0	1	2	3	4+	
0	35	41	18	12	10	116
1	20	42	41	18	10	131
2	13	27	27	09	05	81
3	10	10	11	04	02	37
4+	06	00	05	03	01	15
Total	84	120	102	46	28	380

Table 3: Probabilities of win, draw and loss for each match of the 30^{th} round

Home	Away	Probability			Score	de Finetti	Correct
		Win	Draw	Loss			
Everton	Manchester City	0.705	0.184	0.111	2–0	0.133	Yes
Manchester United	Reading	0.682	0.175	0.143	1–0	0.152	Yes
Aston Villa	Queens Park Rangers	0.612	0.202	0.186	3–2	0.227	Yes
Wigan Athletic	Newcastle United	0.580	0.234	0.186	2–1	0.266	Yes
Stoke City	West Bromwich Albion	0.498	0.255	0.247	0–0	0.863	No
Tottenham Hotspur	Fulham	0.486	0.286	0.228	0–1	0.914	No
Southampton	Liverpool	0.433	0.238	0.329	3–1	0.487	Yes
Chelsea	West Ham United	0.430	0.253	0.318	2–0	0.490	Yes
Sunderland	Norwich City	0.324	0.265	0.411	1–1	0.814	No
Swansea City	Arsenal	0.319	0.277	0.404	0–2	0.535	Yes

3.1. Prediction for a round

In this section, we present the predictions for the 30^{th} round of the EPL. Table 3 shows the probabilities of win, draw and defeat for the 10 games. Table 3 also show the goals scored by each team (score),

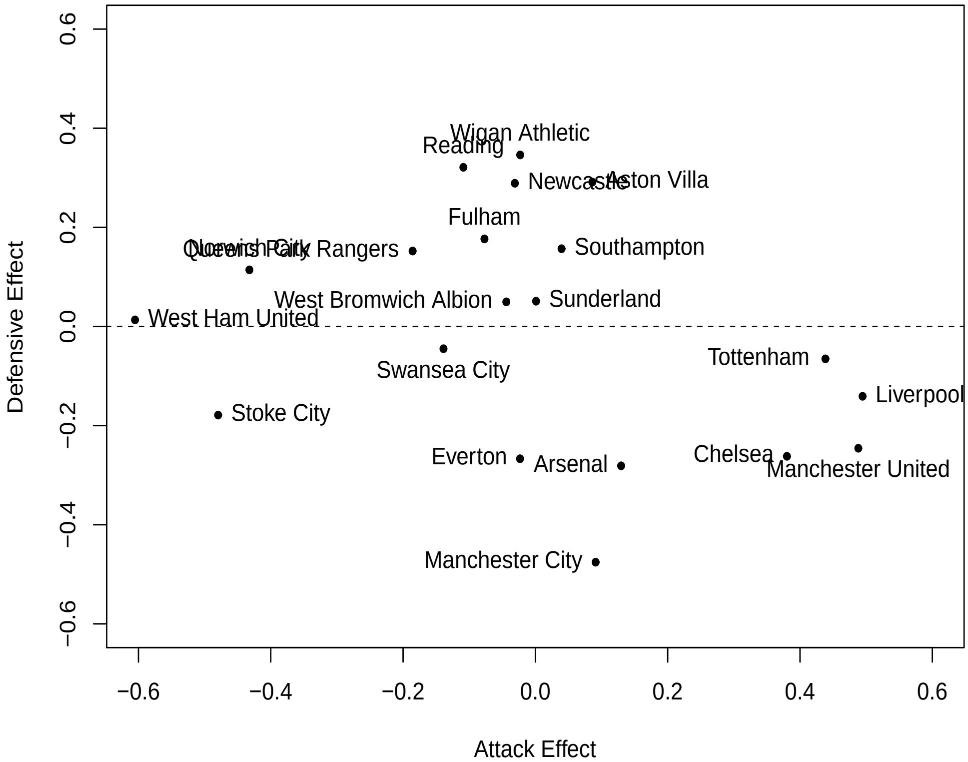

Figure 1: *Attack and defense effect.*

de Finetti measure and if the method correctly indicated the winner team as the team with higher probability of a win. Table 3 is ordered from home team with highest to smallest probability of win.

The proportion of correct prediction was 70%. The teams with an estimated probability of win higher than 0.5, were the actual winning team. If we consider the probability of home team does not loss, i.e., probability of win or tie, then in 80% of games the method correctly indicates the home team as not losing.

Figure 1 displays the graphic of attack versus defense effect. In this graphic, each dot (•) represent the attack and defense effect of each team.

Manchester United and Liverpool have the highest effect attack; while Stoke City and West Ham United have the smallest effect attack. The Manchester City has the best defense effect, i.e., the smallest defense effect; decreasing the expected number of goals of the opposing team.. In opposite, Reading has the worst defense effect, i.e., the highest defense effect, which increases the expected number of goals of the opposing team.

Using a simulation procedure, we estimate the number of points, the number of wins, draw and loss, number of goals for and against for each team. Table 4 presents these values and is organized by the real number of points for each team. The last column in Table 4 show the difference between number of goals for and against. Note that, the six teams with highest estimated number of points are really the six best teams of the championship.

Table 4: Predictions and real values

Team	Points		Won		Drawn		Lost		Goals for		Goals against		Difference of goals	
	Est.	real	Est.	real	Est.	real	Est.	real	Est.	real	Est.	real	Est.	real
Manchester United	92	89	29	28	05	05	04	05	88	86	39	43	49	43
Manchester City	74	78	22	23	08	09	08	06	64	66	39	34	25	32
Chelsea	71	75	20	22	11	09	07	07	75	75	41	39	34	36
Arsenal	71	73	20	21	11	10	07	07	68	66	36	37	32	35
Tottenham Hotspur	65	72	18	21	11	09	09	08	65	66	51	46	14	20
Everton	62	63	16	16	14	15	08	07	57	55	42	40	15	15
Liverpool	56	61	14	16	14	13	10	09	69	71	47	43	22	28
West Bromwich Albion	48	49	14	14	06	07	18	17	48	53	53	57	−05	−04
Swansea City	49	46	12	11	13	13	13	14	50	47	51	51	−01	−04
West Ham United	49	46	13	12	10	10	15	16	45	45	51	53	−06	−08
Norwich City	44	44	10	10	14	14	14	14	38	41	58	58	−20	−17
Fulham	42	43	10	11	12	10	16	17	49	50	61	60	−12	−10
Stoke City	42	42	09	09	15	15	14	14	35	34	46	45	−11	−11
Southampton	44	41	10	09	14	14	14	15	50	49	59	60	−09	−11
Aston Villa	44	41	11	10	11	11	16	17	46	47	66	69	−20	−22
Newcastle United	37	41	10	11	07	08	21	19	44	45	70	68	−26	−23
Sunderland	43	39	11	09	10	12	17	17	46	41	55	54	−09	−13
Wigan Athletic	35	36	09	09	08	09	21	20	43	47	67	73	−24	−26
Reading	32	28	07	06	11	10	20	22	48	43	70	73	−22	−30
Queens Park Rangers	30	25	05	04	15	13	18	21	31	30	57	60	−26	−30

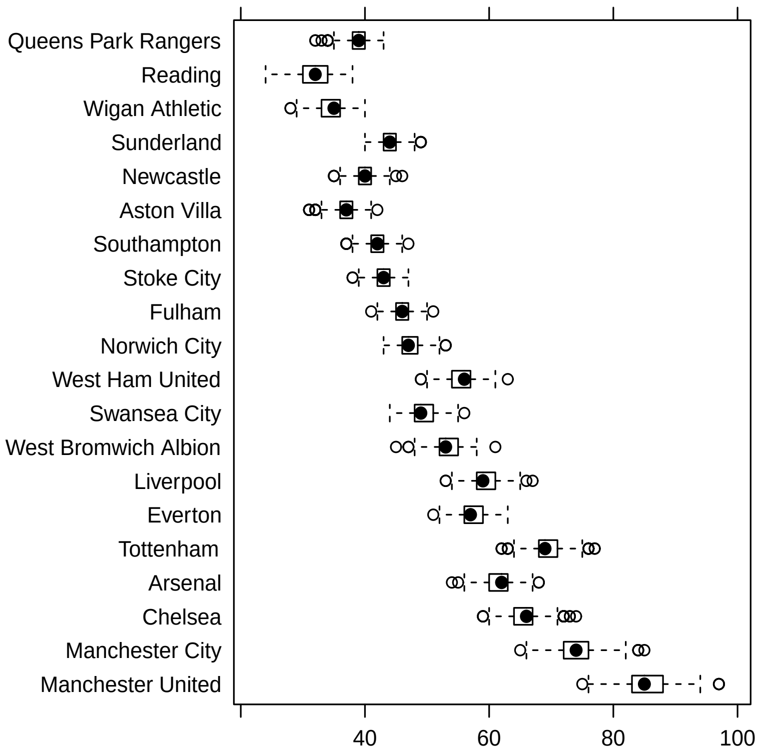

Figure 2: *Box plot of number of points for* $R = 1{,}000$ *simulations.*

Table 5: Probability to be the champion

Rounds	Manchester United	Manchester City	Chelsea	Arsenal	Tottenham Hotspur	Everton	Liverpool
20–38	0.708	0.235	0.025	0.002	0.004	0.008	0.000
22–38	0.662	0.297	0.001	0.000	0.031	0.008	0.001
24–38	0.990	0.007	0.003	0.000	0.000	0.000	0.000
26–38	0.988	0.012	0.000	0.000	0.000	0.000	0.000
28–38	0.873	0.127	0.000	0.000	0.000	0.000	0.000
30–38	0.996	0.003	0.001	0.000	0.000	0.000	0.000
32–38	0.981	0.018	0.001	0.000	0.000	0.000	0.000
34–38	0.998	0.002	0.000	0.000	0.000	0.000	0.000
36–38	1.000	0.000	0.000	0.000	0.000	0.000	0.000
38	1.000	0.000	0.000	0.000	0.000	0.000	0.000

Table 6: Probability to classify for the UEFA Champions League

Rounds	Manchester United	Manchester City	Chelsea	Arsenal	Tottenham	Everton	Liverpool
20–38	0.993	0.955	0.617	0.160	0.337	0.276	0.000
22–38	0.996	0.982	0.402	0.117	0.778	0.506	0.095
24–38	1.000	0.925	0.851	0.203	0.290	0.323	0.257
26–38	1.000	0.992	0.693	0.268	0.799	0.052	0.157
28–38	1.000	0.998	0.823	0.344	0.773	0.012	0.013
30–38	1.000	0.985	0.873	0.148	0.608	0.256	0.100
32–38	1.000	0.998	0.673	0.646	0.543	0.128	0.011
34–38	1.000	1.000	0.697	0.683	0.542	0.078	0.000
36–38	1.000	1.000	0.905	0.527	0.519	0.049	0.000
38	1.000	1.000	1.000	1.000	0.000	0.000	0.000

Table 7: Probability of to be relegated to the second division

Round	Stoke City	Southampton	Aston Villa	Newcastle U.	Sunderland	Wigan A.	Reading	Queens P.R.
20–38	0.000	0.078	0.116	0.250	0.250	0.440	0.813	0.591
22–38	0.004	0.495	0.121	0.053	0.240	0.410	0.621	0.917
24–38	0.010	0.173	0.368	0.100	0.072	0.779	0.663	0.489
26–38	0.007	0.341	0.406	0.001	0.039	0.864	0.643	0.580
28–38	0.034	0.095	0.719	0.093	0.015	0.342	0.711	0.812
30–38	0.008	0.323	0.314	0.063	0.076	0.484	0.796	0.869
32–38	0.007	0.020	0.213	0.030	0.199	0.624	0.874	0.943
34–38	0.153	0.004	0.217	0.028	0.119	0.432	0.986	0.999
36–38	0.000	0.001	0.102	0.115	0.027	0.720	1.000	1.000
38	0.000	0.000	0.000	0.000	0.000	1.000	1.000	1.000

Figure 2 shows box plot of the estimated points for the 20 teams from $R = 1{,}000$ simulations. As we can note, the simulation results show Manchester United as champion.

3.2. Predictions for whole second phase

We apply the proposed method to predict results of the matches of rounds 20 to 38. Using a simulation procedure, we calculated the probability of each team being the champion and to classify the Union of European Football Associations (UEFA) champions league.

Tables 5–7 below show results for rounds $20, 22, 24, \ldots, 38$. Tables B.1–B.3 in Appendix B of the SM show results for all rounds. Table 5 shows the rounds simulated and the probabilities to be champion for the seven teams with the highest number of goals scored (Table 1). In all second phase of the champion the method indicates Manchester United as the champion with a probability higher

Table 8: Number of games and goals scored by each team at home and away in each phase of the BFL

Team	1st phase				2nd phase				Number of goals		
	Games at home	Goals	Games away	Goals	Games at home	Goals	Games away	Goals	At home	Away	Total
Corinthians	09	17	10	10	10	24	09	20	41	30	71
Atlético-MG	10	19	09	14	09	17	10	15	36	29	65
Palmeiras	10	19	09	13	09	12	10	16	31	29	60
Santos	09	16	10	09	10	31	09	03	47	12	59
São Paulo	10	16	09	09	09	19	10	09	35	18	53
Sport	10	18	09	13	09	15	10	07	33	20	53
Grêmio	10	21	09	08	09	14	10	09	35	17	52
Flamengo	09	12	10	09	10	16	09	08	28	17	45
Cruzeiro	09	08	10	07	10	20	09	09	28	16	44
Atlético Paranaense	10	14	09	09	09	16	10	04	30	13	43
Ponte Preta	09	09	10	12	10	13	09	07	22	19	41
Fluminense	10	12	09	10	09	13	10	05	25	15	40
Internacional	09	09	10	05	10	19	09	06	28	11	39
Goiás	09	08	10	08	10	14	09	09	22	17	39
Avaí	10	13	09	05	09	13	10	07	26	12	38
Figueirense	09	10	10	08	10	08	09	10	18	18	36
Chapecoense	10	14	09	03	09	09	10	08	23	11	34
Coritiba	09	08	10	05	10	07	09	11	15	16	31
Vasco da Gama	10	05	09	03	09	08	10	12	13	15	28
Joinvile	09	09	10	04	10	10	09	03	19	07	26

than 0.63. After the 28^{th} round, the probability of Manchester United be the champion is higher than 0.99.

Table 6 shows the probabilities for the seven teams with highest number of goals to classify for the UEFA Champions League. In all second phase, Manchester United has a probability to classify to the UEFA Champions League that is higher than 0.99. After 28^{th} round the probability of Manchester City to classify is higher than 0.99. Six rounds before the ending of the champion, the method indicates Manchester United and Manchester City as the two teams classified for the UEFA Champions League.

Table 7 shows the probabilities of the eight teams with smallest estimated number of points to be relegated to the second division (Table 4). In the 35^{th} round (four rounds before the end of the championship), the method indicates these both teams as the teams relegated to the second division.

Figures C.1 and C.2 in Appendix C of the SM shows the attack effect and defense effect for the best four teams of the EPL in the 20–38 rounds.

4. Application 2

In this section we apply the proposed method to the BFL. As EPL, the BFL it is also composed by $n + 1 = 20$ teams. The games are played in two phases, in which each team plays 19 games by phase. At the end of the 38 rounds, the team with highest number of points is champion; the four teams with the highest number of points are classified to 2016 Copa Libertadores of América and the four teams with the smallest points are relegated to the second division of the BFL.

Table 8 shows the number of games and the number of goals scored by each team at home and away in each phase of the BFL. The two best team of BFL, Corinthians and Atlético-MG, have the highest number of goals scored; being that the best team, Corinthians, has the highest number of goals scored at home and away. The two worst teams of the BFL, Joinvile and Vasco, are the teams with smallest number of goals scored. Joinvile has the smallest number of goals scored away, while Vasco

Table 9: Number of games ended with number of goals (x_t, x_s), for $x_t, x_s \in \{0, 1, 2, 3, +4\}$

Away team	Home team					Total
	0	1	2	3	+4	
0	39	56	35	28	4	162
1	27	33	43	15	10	128
2	16	21	16	5	3	61
3	5	11	4	3	1	24
+4	1	4	0	0	0	5
Total	88	125	98	51	18	380

Table 10: Probabilities of win, draw and loss for each match of the 27^{th} round

Home	Away	Probability			Score	de Finetti	Correct
		Win	Draw	Loss			
Palmeiras	Grêmio	0.590	0.256	0.154	3–2	0.258	Yes
Internacional	Figueirense	0.324	0.228	0.448	1–1	0.903	No
Ponte Preta	Fluminense	0.499	0.233	0.268	3–1	0.377	Yes
Corinthians	Santos	0.576	0.208	0.216	2–0	0.270	Yes
Goiás	Joinvile	0.766	0.136	0.098	3–0	0.083	Yes
Vasco	Sport	0.607	0.222	0.172	2–1	0.233	Yes
Atlético-MG	Flamengo	0.573	0.226	0.201	4–1	0.273	Yes
Avaí	São Paulo	0.337	0.332	0.331	2–1	0.659	Yes
Coritiba	Atlético-PR	0.387	0.292	0.321	2–0	0.564	Yes
Chapecoense	Cruzeiro	0.484	0.262	0.254	0–2	0.859	No

has the smallest number of goals scored at home.

Table 9 presents the number of games that end with the number of goals (x_t, x_s). In 52.63% of the games the home team was the winner; in 23.42% the winner was the away team and in 23.95% the game ended in a draw. The amount of victory of the home team is more than twice the amount of victories for the away team. The two most frequent results was 1–0 (56 games) and 2–1 (43 games) for the home team.

4.1. Prediction for a round

Here we present the predictions of the 27^{th} round of the BFL. Table 10 presents the probabilities of win, draw, goals scored, de Finetti measure and if the method indicates the winning team as the team with higher probability of win. The percentage of correct predictions was 80%.

Figure 3 displays the graphic of attack versus defense effect. Palmeiras, Atlético-MG and Corinthians have the highest attack effect. Table 10 shows that these teams won their games. Chapecoense, Internacional and Joinvile have the worst attack effect. Corinthians also presents the better defense effect, while Vasco, Avaí and Figueirense have the worst defense effect.

Using the simulation procedure, we estimate the number of points for each team. Figure 4 shows box plot of the estimated points for the 20 teams of the BFL from $R = 1{,}000$ simulations. As we can note, the simulation results show Corinthians as the champion. Corinthians, Atlético-MG, Grêmio and São Paulo as teams classified for 2016 Copa Libertadores of América; and Vasco, Joinvile, Figueirense and Goiás as teams relegated to the second division. From real results, the method correctly indicates the champion, the four teams classified for 2016 Copa Libertadores of América; and the three teams relegated to the second division: Vasco, Joinvile and Goiás. The fourth team relegated to the second division was Avaí and not Figueirense as foreseen by the method.

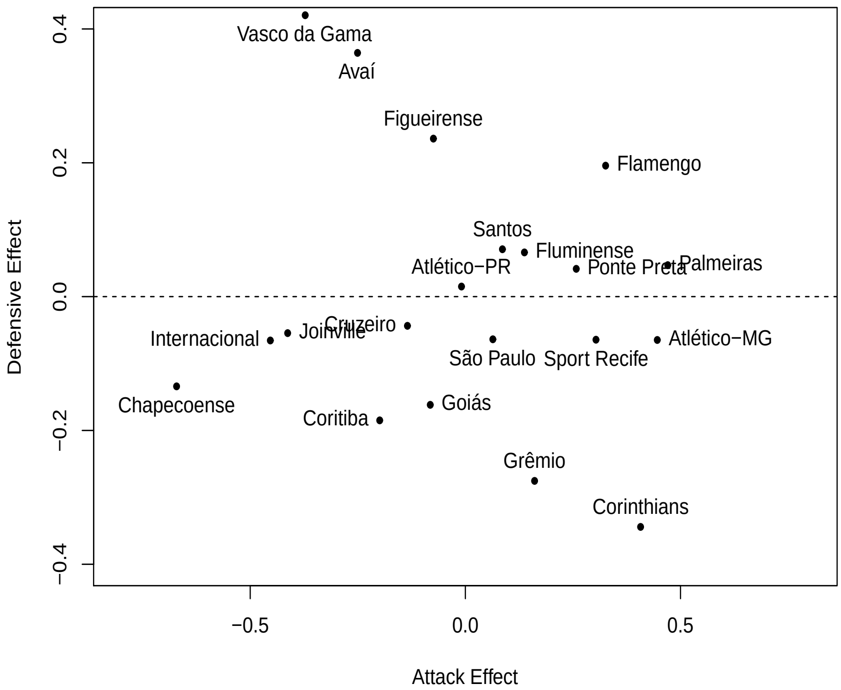

Figure 3: *Attack and defense effect.*

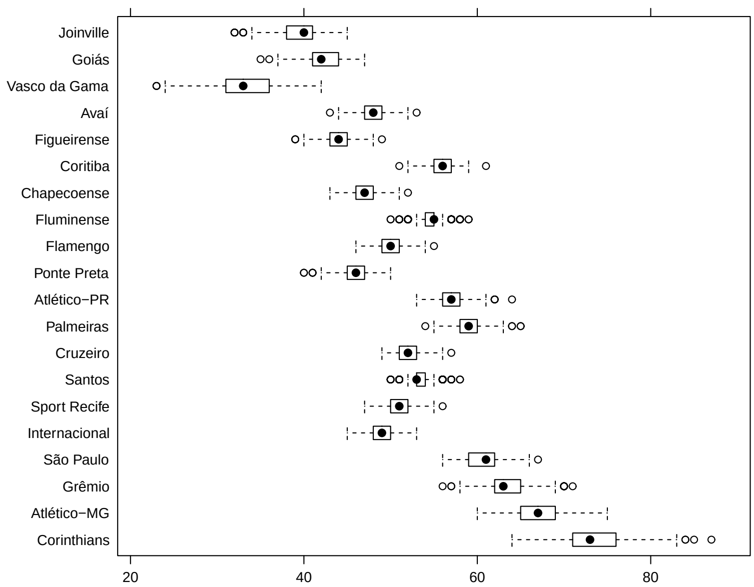

Figure 4: *Box plot of number of points for* $R = 1{,}000$ *simulations.*

Table 11: Probability to be the champion

Rounds	Corinthians	Atlético-MG	Grêmio	São Paulo	Internacional	Sport Recife	Santos
20–38	0.572	0.047	0.065	0.125	0.000	0.008	0.000
22–38	0.176	0.204	0.185	0.007	0.031	0.028	0.002
24–38	0.692	0.086	0.194	0.020	0.000	0.006	0.000
26–38	0.579	0.380	0.025	0.010	0.000	0.000	0.002
28–38	0.945	0.035	0.007	0.002	0.000	0.002	0.001
30–38	0.728	0.230	0.012	0.020	0.000	0.000	0.007
32–38	0.940	0.058	0.002	0.000	0.000	0.000	0.000
34–38	0.999	0.001	0.000	0.000	0.000	0.000	0.000
36–38	1.000	0.000	0.000	0.000	0.000	0.000	0.000
38	1.000	0.000	0.000	0.000	0.000	0.000	0.000

Table 12: Probability to classify for the 2016 Copa Libertadores of América

Rounds	Corinthians	Atlético-MG	Grêmio	São Paulo	Internacional	Sport Recife	Santos
20–38	0.953	0.467	0.487	0.684	0.000	0.135	0.003
22–38	0.606	0.676	0.573	0.048	0.231	0.178	0.030
24–38	0.991	0.856	0.943	0.548	0.032	0.274	0.011
26–38	0.992	0.981	0.712	0.399	0.025	0.010	0.164
28–38	0.999	0.751	0.540	0.231	0.100	0.251	0.204
30–38	0.996	0.942	0.456	0.587	0.101	0.002	0.345
32–38	1.000	0.997	0.927	0.330	0.131	0.034	0.203
34–38	1.000	1.000	0.971	0.140	0.123	0.200	0.460
36–38	1.000	1.000	1.000	0.488	0.014	0.037	0.390
38	1.000	1.000	1.000	0.859	0.111	0.030	0.000

Table 13: Probability of to be relegated to the second division

Round	Joinville	Goiás	Vasco da Gama	Avaí	Figueirense	Coritiba
20–38	0.381	0.330	0.301	0.513	0.317	0.798
22–38	0.462	0.703	0.981	0.535	0.103	0.306
24–38	0.782	0.509	0.995	0.664	0.111	0.151
26–38	0.800	0.155	0.996	0.272	0.264	0.208
28–38	0.882	0.378	0.987	0.506	0.662	0.084
30–38	0.978	0.449	0.869	0.120	0.491	0.635
32–38	0.645	0.646	0.954	0.799	0.254	0.601
34–38	0.977	0.407	0.982	0.625	0.081	0.830
36–38	0.999	0.915	0.971	0.443	0.086	0.581
38	1.000	0.949	0.911	0.421	0.711	0.008

4.2. Predictions for whole second phase

In this section, we apply the proposed method to predict results of the matches of rounds 20 to 38. Table 11 shows the rounds simulated and the probabilities to be champion for the seven teams with the highest estimated number of points. After 31-round the method indicates Corinthians as the champion with a probability higher than 0.93. The method indicates the Corinthians as the champion three rounds before the ending.

Table 12 shows the probabilities for the seven teams with the highest estimated number of points to classify for the 2016 Copa Libertadores of América. Eight rounds before the ending of the BFL, the probability of the Corinthians and Atlético-MG to classify for the 2016 Copa Libertadores of América is higher than 0.990. Three rounds before the ending of the champion, the method indicates the Corinthians, Atlético-MG and Grêmio as teams classified 2016 Copa Libertadores of América. In the last round, the method indicates São Paulo as the fourth team classified with the probability 0.859.

Table 13 shows the probabilities of the six teams with smallest estimated number of points to be relegated to the second division. Three rounds before the ending of the champion, the method indicates the Joinvile, Goiás and Vasco as teams relegated to the second division with a probability higher than 0.90. In the last round, the method indicates Avaí and Figueirense with probabilities 0.421 and 0.711 to be relegated. The Avaí was relegated.

Tables D.1–D.3 in Appendix D of the show results for all rounds. Figures E.1 and E.2 in Appendix E of the SM show, respectively, the attack effect and defense effect for the best four teams and worst four teams of the BFL in the 20–38 rounds.

5. Final remarks

In this paper, we develop a model to estimate the probabilities of win, tie and defeat in football games. In order to calculate these probabilities we propose a Poisson regression model, in which, the average of goals scored reflects the strength of attack of the team, the strength of defense of the opposing team and the home team effect. Inferences on parameters of interest were done via Bayesian inference. The accuracy of the forecasts were measured using the de Finetti measure.

In order to illustrate the application of the proposed method, we apply it to the 2012–2013 EPL and to 2015 BFL. Using a simulation procedure, we calculated satisfactory results on the probability of each team being the champion and classify the continental tournaments. The method correctly indicated the champion of the EPL and BFL with three rounds before the ending of the championship. The method also correctly indicates the teams classified for the continental tournaments.

We also present the probabilities of the teams to be relegated to the second division. Again, the method presents satisfactorily results. The attack effect and defense effect for the best four teams and worst four teams of the EPL and BFL in the 20–38 rounds were also presented. Results showed that champions team have a higher attack effect and smaller defense effect. These two facts, increase in the expected number of goals of the champion team and decrease the expected number of goals of an opposing team, respectively.

Results obtained show that proposed method may be an effective alternative to predict football outcomes. A practical differential of the proposed method is its simplicity to be implemented in software like OpenBUGS and R.

We have developed a Poisson regression model based on the average of goals scored that reflect the strength of attack and defense of the teams; in addition, the model can also describe the means of goals scored using other covariates, such as, atmospheric condition, injuries, suspensions, tactical scheme, and crisis. It can be seems as future work. Besides, the method can be easily used for the analysis of the upcoming championship season and adapted to other championship and tournaments with different forms of dispute.

All computational were performed using OpenBUGS and R systems via R2WinBUGS package. The computer programs are also available in the Supplementary Material at CSAM homepage (http://csam.or.kr).

Acknowledgements

We thank the Editor and the referees for their comments, suggestions and criticisms which have led to improvements in this article. The first author acknowledges the Brazilian institution CNPq. The research of Francisco Louzada is supported by CNPq and FAPESP.

Appendix

Here we provide some additional results and the computational codes developed to calculate the probabilities presented in the manuscript.

Appendix A: Estimations of βs

In this appendix we present some details of the estimation procedure of parameters $\boldsymbol{\beta}_k = (\beta_{k0}, \beta_{ka}, \beta_{kd}, \beta_{kh})$, for $k = t, s$.

From model (2.1) of the manuscript, the likelihood function is given by

$$L(\boldsymbol{\beta}_t, \boldsymbol{\beta}_s | \mathbf{x}_t, \mathbf{x_s}) = \prod_{k \in \{t,s\}} \prod_{m=1}^{n_r} \frac{e^{x_{km} U_{km} \boldsymbol{\beta}_k} e^{-e^{U_{km}\boldsymbol{\beta}_k}}}{x_{km}!}. \tag{A.1}$$

The log-likelihood function is given in equation (2.3) of the manuscript.

We assume that priors are *a priori* independent, i.e., $\pi(\boldsymbol{\beta}_t, \boldsymbol{\beta}_s) = \pi(\boldsymbol{\beta}_t)\pi(\boldsymbol{\beta}_s)$, in which, $\pi(\boldsymbol{\beta}_k) = \pi(\beta_{k0})\pi(\beta_{ka})\pi(\beta_{kd})\pi(\beta_{kh})$, for $k = t, s$. So, we consider the following prior distributions:

$$\beta_{k0} \sim \mathcal{N}\left(0, 10^{-4}\right), \quad \beta_{ka} \sim \mathcal{N}\left(0, 10^{-3}\right), \quad \beta_{k^c d} \sim \mathcal{N}\left(0, 10^{-3}\right) \quad \text{and} \quad \beta_{kh} \sim \mathcal{N}\left(0, 10^{-3}\right),$$

for $k = t, s$, where $\mathcal{N}(0, b)$ denotes the normal distribution with mean 0 and precision b.

Here, we present the joint posterior distribution for parameters $\boldsymbol{\beta}_t$. The joint posterior distribution for $\boldsymbol{\beta}_s$ is obtained in a similar way.

Updating the joint prior distribution for $\pi(\boldsymbol{\beta}_t)$ via likelihood function in (A.1), the joint posterior distribution is given by

$$\pi(\boldsymbol{\beta}_t | \mathbf{x_t}) \propto \left[\prod_{m=1}^{n_r} e^{x_{tm} U_{tm} \boldsymbol{\beta}_t} e^{-e^{U_{tm}\boldsymbol{\beta}_t}}\right] \pi(\beta_{t0})\pi(\beta_{ta})\pi(\beta_{sd})\pi(\beta_{th}). \tag{A.2}$$

The conditional posterior distributions for β_{tw}, $w \in \{0, a, d, h\}$, is given by

$$\pi(\beta_{tw} | \mathbf{x_t}, \boldsymbol{\beta}_t \setminus \beta_{tw}) \propto \left[\prod_{m=1}^{n_r} e^{x_{tm} U_{tm} \boldsymbol{\beta}_t} e^{-e^{U_{tm}\boldsymbol{\beta}_t}}\right] \pi(\beta_{tw}),$$

where $\boldsymbol{\beta}_t \setminus \beta_{tw}$ denotes the vector $\boldsymbol{\beta}_t$ excluding β_{tw}.

As one can note, the conditional posterior distribution for β_{t0}, β_{ta}, β_{sd} and β_{th} do not follow any close distribution. For this case, the usual Bayesian procedure to generate random samples from posterior distribution is to use the Metropolis-Hastings (MH) algorithm.

The Metropolis-Hastings algorithm together with the Gibbs sampling are the two most popular examples of a Markov chain Monte Carlo (MCMC) method. This algorithm is used for sampling from generic distributions if we do not know how to generate a random sample. Similar to acceptance-rejection sampling, the MH algorithm considers that (to each iteration of the algorithm) a candidate value can be generated from a proposal density; therefore, the candidate value is accepted according to an adequated acceptance probability. This procedure guarantees the convergency of the Markov chain for the target density. For more details on MH algorithm see Hastings (1970), Chib and Greenberg (1995), Gelman *et al.* (1995) and Gilks *et al.* (1996).

Table B.1: Probability to be the champion

Rounds	Manchester United	Manchester City	Chelsea	Arsenal	Tottenham Hotspur	Everton	Liverpool
20–38	0.708	0.235	0.025	0.002	0.004	0.008	0.000
21–38	0.815	0.031	0.129	0.001	0.020	0.000	0.000
22–38	0.662	0.297	0.001	0.000	0.031	0.008	0.001
23–38	0.925	0.015	0.001	0.000	0.045	0.004	0.000
24–38	0.990	0.007	0.003	0.000	0.000	0.000	0.000
25–38	0.633	0.345	0.008	0.000	0.005	0.009	0.000
26–38	0.988	0.012	0.000	0.000	0.000	0.000	0.000
27–38	0.972	0.006	0.009	0.000	0.013	0.000	0.000
28–38	0.873	0.127	0.000	0.000	0.000	0.000	0.000
29–38	0.995	0.004	0.000	0.000	0.001	0.000	0.000
30–38	0.996	0.003	0.001	0.000	0.000	0.000	0.000
31–38	0.993	0.007	0.000	0.000	0.000	0.000	0.000
32–38	0.981	0.018	0.001	0.000	0.000	0.000	0.000
33–38	0.989	0.011	0.000	0.000	0.000	0.000	0.000
34–38	0.998	0.002	0.000	0.000	0.000	0.000	0.000
35–38	1.000	0.000	0.000	0.000	0.000	0.000	0.000
36–38	1.000	0.000	0.000	0.000	0.000	0.000	0.000
37–38	1.000	0.000	0.000	0.000	0.000	0.000	0.000
38	1.000	0.000	0.000	0.000	0.000	0.000	0.000

For example, to update parameter β_{th} via MH algorithm, consider $(\beta_{t0}, \beta_{ta}, \beta_{sd}, \beta_{th})$ be the current state of the Markov chain. Let β^*_{th} be a candidate value generated from a candidate generating-density distribution $q[\beta^*_{th}|\beta_{th}]$. So, the value β^*_{th} is accepted with probability $\Psi(\beta^*_{th}|\beta_{th}) = \min(1, A_{\beta_{th}})$, where

$$A_{\beta_{th}} = \frac{L\left(\beta_{t0}, \beta_{ta}, \beta_{sd}, \beta^*_{th}|\mathbf{x}_t\right)\pi\left(\beta^*_{th}\right)}{L\left(\beta_{t0}, \beta_{ta}, \beta_{sd}, \beta_{th}|\mathbf{x}_t\right)\pi\left(\beta_{th}\right)} \frac{q\left[\beta_{th}|\beta^*_{th}\right]}{q\left[\beta^*_{th}|\beta_{th}\right]} \tag{A.3}$$

and $L(\beta_{t0}, \beta_{ta}, \beta_{sd}, \beta_{th}|\mathbf{x}_t) \propto [\prod_{m=1}^{n_r} e^{x_{tm}U_{tm}\boldsymbol{\beta}_t} e^{-e^{U_{tm}\boldsymbol{\beta}_t}}]$ is the likelihood function for $\boldsymbol{\beta}_t$.

In practical terms, the MH algorithm is implemented as follows.

Metropolis-Hastings algorithm: Let the current state of the Markov chain consist of $(\beta^{(l)}_{t0}, \beta^{(l)}_{ta}, \beta^{(l)}_{sd}, \beta^{(l-1)}_{th})$, where l is l^{th} iteration of the algorithm, for $l = 1, \dots, L$. So, update β_{th} as follows:

(1) Generate $\beta^*_{th} \sim q[\beta^*_{th}|\beta_{th}]$;

(2) Calculate $\Psi(\beta^*_{th}|\beta_{th}) = \min(1, A_{\beta_{th}})$, where $A_{\beta_{th}}$ is given in (A.3);

(3) Generate $u \sim U(0, 1)$. If $u \leq \Psi(\beta^*_{th}|\beta_t h)$ accept β^*_{th} and do $\beta^{(l)}_{th} = \beta^*_{th}$. Otherwise, reject β^*_{th} and do $\beta^{(l)}_{th} = \beta^{(l-1)}_{th}$.

The procedure to update β_{t0}, β_{ta} and β_{sd} is similar to described to parameter β_{th}. We implement the MH in order to generate random samples from posterior distribution in (A.2) using the WinBUGS (Spiegelhalter *et al.*, 2003) and R (R Development Core Team, 2012) softwares via R2WinBUGS package Gelman *et al.* (2006). The computer programs are available in the Supplementary Material at CSAM homepage (http://csam.or.kr).

Appendix B: Estimated probabilities for EPL

In this appendix we present the complete version of the Tables 5–7 showed in manuscript. Tables B.1 and B.2 show all rounds simulated, the probabilities to be champion and to classify for the continental

Table B.2: Probability to classify for the UEFA Champions League

Rounds	Manchester United	Manchester City	Chelsea	Arsenal	Tottenham	Everton	Liverpool
20–38	0.993	0.955	0.617	0.160	0.337	0.276	0.000
21–38	0.998	0.701	0.918	0.081	0.591	0.213	0.034
22–38	0.996	0.982	0.402	0.117	0.778	0.506	0.095
23–38	0.998	0.724	0.297	0.153	0.896	0.284	0.017
24–38	1.000	0.925	0.851	0.203	0.290	0.323	0.257
25–38	1.000	0.993	0.528	0.046	0.547	0.552	0.297
26–38	1.000	0.992	0.693	0.268	0.799	0.052	0.157
27–38	1.000	0.848	0.880	0.189	0.905	0.076	0.009
28–38	1.000	0.998	0.823	0.344	0.773	0.012	0.013
29–38	1.000	0.995	0.802	0.033	0.776	0.159	0.214
30–38	1.000	0.985	0.873	0.148	0.608	0.256	0.100
31–31	1.000	0.975	0.927	0.170	0.774	0.079	0.069
32–38	1.000	0.998	0.673	0.646	0.543	0.128	0.011
33–38	1.000	1.000	0.965	0.332	0.655	0.037	0.011
34–38	1.000	1.000	0.697	0.683	0.542	0.078	0.000
35–38	1.000	1.000	0.810	0.737	0.425	0.028	0.000
36–38	1.000	1.000	0.905	0.527	0.519	0.049	0.000
37–38	1.000	1.000	0.961	0.775	0.264	0.000	0.000
38	1.000	1.000	1.000	1.000	0.000	0.000	0.000

Table B.3: Probability of to be relegated to the second division

Round	Stoke City	Southampton	Aston Villa	Newcastle United	Sunderland	Wigan Athletic	Reading	Queens P.R.
20–38	0.000	0.078	0.116	0.250	0.250	0.440	0.813	0.591
21–38	0.006	0.308	0.042	0.262	0.294	0.073	0.863	0.754
22–38	0.004	0.495	0.121	0.053	0.240	0.410	0.621	0.917
23–38	0.068	0.015	0.386	0.597	0.121	0.214	0.976	0.162
24–38	0.010	0.173	0.368	0.100	0.072	0.779	0.663	0.489
25–38	0.005	0.345	0.456	0.387	0.084	0.495	0.178	0.846
26–38	0.007	0.341	0.406	0.001	0.039	0.864	0.643	0.580
27–38	0.000	0.076	0.165	0.666	0.180	0.339	0.484	0.855
28–38	0.034	0.095	0.719	0.093	0.015	0.342	0.711	0.812
29–38	0.011	0.216	0.867	0.147	0.016	0.338	0.570	0.751
30–38	0.008	0.323	0.314	0.063	0.076	0.484	0.796	0.869
31–38	0.021	0.201	0.252	0.018	0.127	0.439	0.923	0.968
32–38	0.007	0.020	0.213	0.030	0.199	0.624	0.874	0.943
33–38	0.070	0.003	0.250	0.010	0.291	0.373	0.985	0.975
34–38	0.153	0.004	0.217	0.028	0.119	0.432	0.986	0.999
35–38	0.014	0.003	0.290	0.040	0.072	0.554	1.000	1.000
36–38	0.000	0.001	0.102	0.115	0.027	0.720	1.000	1.000
37–38	0.000	0.005	0.004	0.131	0.024	0.620	1.000	1.000
38	0.000	0.000	0.000	0.000	0.000	1.000	1.000	1.000

championship for the seven teams with the highest number of goals scored. Table B.3 shows all rounds simulated and the probabilities of the eight teams with smallest estimated number of points to be relegated to the second division.

Appendix C: Attack and defense effect for EPL

In this appendix we present the attack and defense effect for the best four teams and worst four teams of the EPL and BFL in the 20–38 rounds.

Figure C.1 shows the attack effect and defense effect for the best four teams of the EPL in the

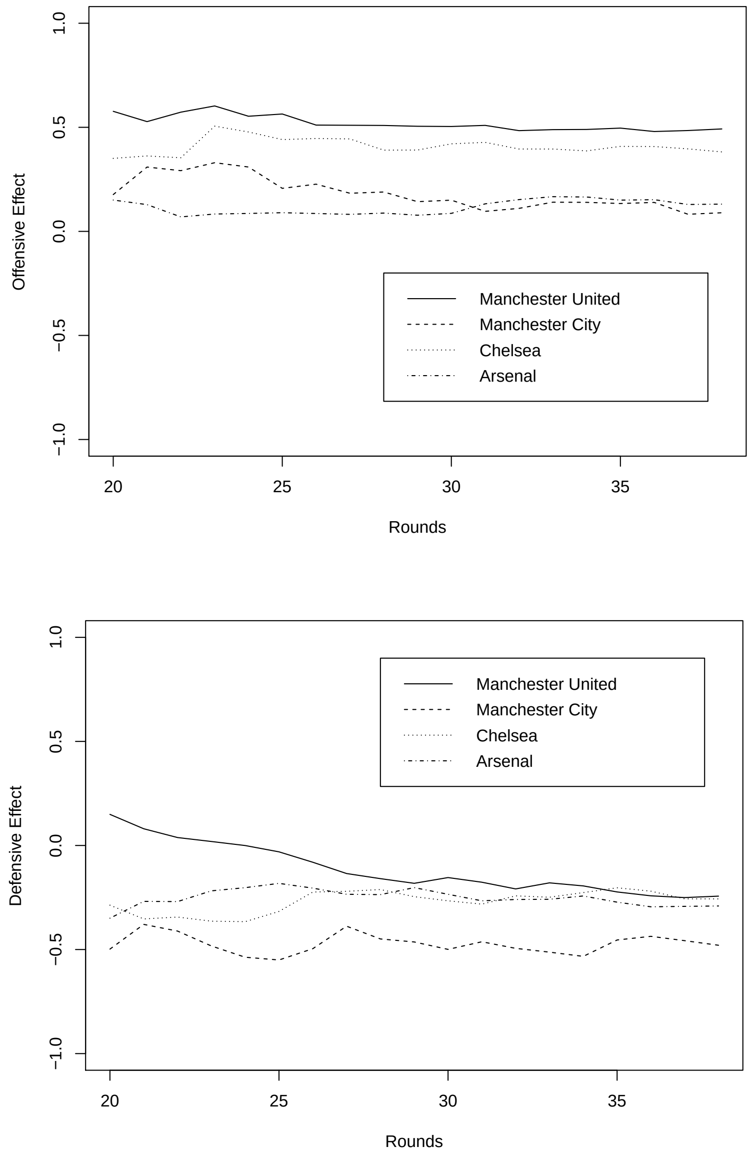

Figure C.1: *Attack and defense effect of the best four teams in rounds 20–38.*

20–38 rounds. These four teams present positive attack. Manchester City, Chelsea and Arsenal have negative defense affect in the 20–38 rounds. The positive attack increases the expected number of goals of the teams and the negative defense attack decrease the expected number of goals of the opposing team.

In opposite, the two worst team of the EPL, Queens Park Rangers and Reading have negative attack effect and positive defense attack, as showed in Figure C.2. This Figure C.2 also show the

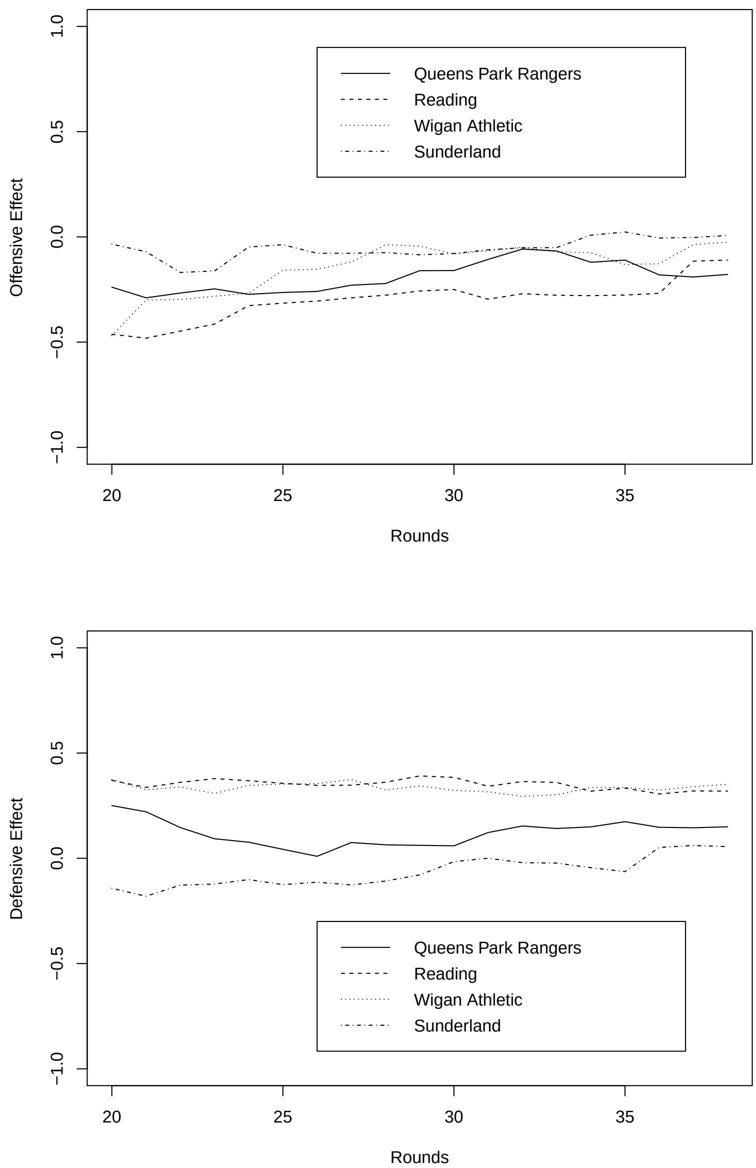

Figure C.2: *Attack and defense effect of the worst four teams in rounds 20–38.*

attack and defense effect for Wigan Athletic and Sunderland.

Appendix D: Estimated probabilities for BFL

In this appendix we present the complete version of the Tables 11–13 showed in manuscript. Tables D.1 and D.2 show all rounds simulated, the probabilities to be champion and to classify for the

Table D.1: Probability to be the champion

Rounds simulated	Corinthians	Atlético-MG	Grêmio	São Paulo	Internacional	Sport Recife	Santos
20–38	0.572	0.047	0.065	0.125	0.000	0.008	0.000
21–38	0.343	0.298	0.214	0.073	0.000	0.010	0.030
22–38	0.176	0.204	0.185	0.007	0.031	0.028	0.002
23–38	0.465	0.199	0.290	0.016	0.000	0.001	0.015
24–38	0.692	0.086	0.194	0.020	0.000	0.006	0.000
25–38	0.386	0.388	0.099	0.002	0.016	0.000	0.000
26–38	0.579	0.380	0.025	0.010	0.000	0.000	0.002
27–38	0.430	0.505	0.048	0.002	0.005	0.002	0.004
28–38	0.945	0.035	0.007	0.002	0.000	0.002	0.001
29–38	0.953	0.040	0.007	0.000	0.000	0.000	0.000
30–38	0.728	0.230	0.012	0.020	0.000	0.000	0.007
31–38	0.439	0.394	0.166	0.000	0.000	0.000	0.000
32–38	0.940	0.058	0.002	0.000	0.000	0.000	0.000
33–38	0.953	0.047	0.000	0.000	0.000	0.000	0.000
34–38	0.999	0.001	0.000	0.000	0.000	0.000	0.000
35–38	0.999	0.001	0.000	0.000	0.000	0.000	0.000
36–38	1.000	0.000	0.000	0.000	0.000	0.000	0.000
37–38	1.000	0.000	0.000	0.000	0.000	0.000	0.000
38	1.000	0.000	0.000	0.000	0.000	0.000	0.000

Table D.2: Probability to classify for the 2016 Copa Libertadores of América

Rounds simulated	Corinthians	Atlético-MG	Grêmio	São Paulo	Internacional	Sport Recife	Santos
20–38	0.953	0.467	0.487	0.684	0.000	0.135	0.003
21–38	0.865	0.821	0.765	0.441	0.001	0.206	0.359
22–38	0.606	0.676	0.573	0.048	0.231	0.178	0.030
23–38	0.951	0.875	0.884	0.272	0.039	0.054	0.296
24–38	0.991	0.856	0.943	0.548	0.032	0.274	0.011
25–38	0.936	0.926	0.683	0.117	0.299	0.008	0.008
26–38	0.992	0.981	0.712	0.399	0.025	0.010	0.164
27–38	0.983	0.981	0.709	0.156	0.255	0.114	0.322
28–38	0.999	0.751	0.540	0.231	0.100	0.251	0.204
29–38	1.000	0.969	0.894	0.105	0.011	0.067	0.237
30–38	0.996	0.942	0.456	0.587	0.101	0.002	0.345
31–38	0.999	0.999	0.990	0.173	0.029	0.045	0.296
32–38	1.000	0.997	0.927	0.330	0.131	0.034	0.203
33–38	1.000	0.999	0.873	0.097	0.184	0.417	0.396
34–38	1.000	1.000	0.971	0.140	0.123	0.200	0.460
35–38	1.000	1.000	0.962	0.213	0.064	0.183	0.547
36–38	1.000	1.000	1.000	0.488	0.014	0.037	0.390
37–38	1.000	1.000	1.000	0.314	0.367	0.008	0.281
38	1.000	1.000	1.000	0.859	0.111	0.030	0.000

continental championship for the seven teams with the highest number of goals scored. Table D.3 shows the probabilities of the six teams with smallest estimated number of points to be relegated to the second division.

Appendix E: Attack and defense effect for BFL

In this appendix we present the attack and defense effect for the best four and worst four teams of the BFL in the 20–38 rounds.

Figures E.1 and E.2 show, respectively, the attack effect and defense effect for the best four teams

Table D.3: Probability of to be relegated to the second division

Round simulated	Joinville	Goiás	Vasco da Gama	Avaí	Figueirense	Coritiba
20–38	0.381	0.330	0.301	0.513	0.317	0.798
21–38	0.923	0.528	0.983	0.053	0.628	0.217
22–38	0.462	0.703	0.981	0.535	0.103	0.306
23–38	0.842	0.262	0.910	0.629	0.369	0.416
24–38	0.782	0.509	0.995	0.664	0.111	0.151
25–38	0.701	0.152	0.728	0.536	0.174	0.553
26–38	0.800	0.155	0.996	0.272	0.264	0.208
27–38	0.940	0.337	0.948	0.419	0.802	0.155
28–38	0.882	0.378	0.987	0.506	0.662	0.084
29–38	0.600	0.726	0.917	0.547	0.618	0.124
30–38	0.978	0.449	0.869	0.120	0.491	0.635
31–38	0.904	0.565	0.944	0.274	0.509	0.441
32–38	0.645	0.646	0.954	0.799	0.254	0.601
33–38	0.859	0.762	0.942	0.317	0.505	0.431
34–38	0.977	0.407	0.982	0.625	0.081	0.830
35–38	0.909	0.626	0.856	0.701	0.169	0.739
36–38	0.999	0.915	0.971	0.443	0.086	0.581
37–38	1.000	0.953	0.864	0.651	0.246	0.286
38	1.000	0.949	0.911	0.421	0.711	0.008

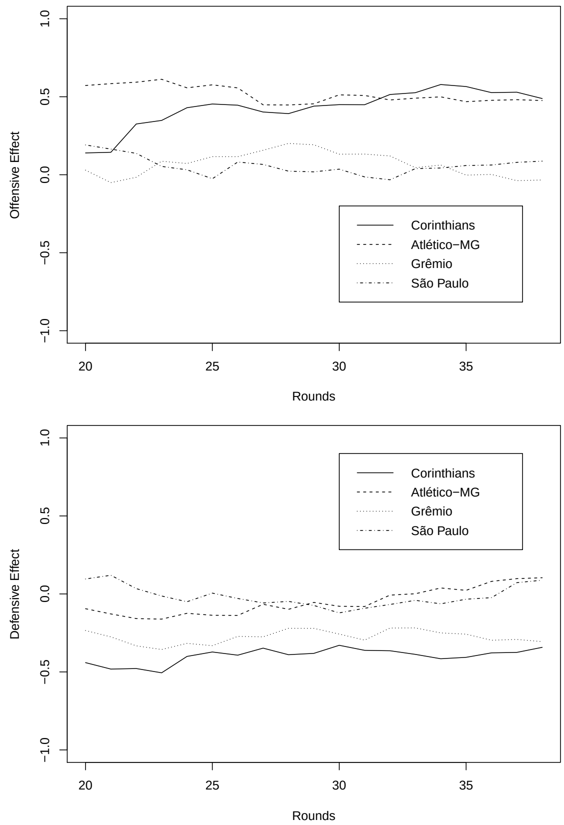

Figure E.1: *Attack and defense effect of the best four teams in rounds 20–38.*

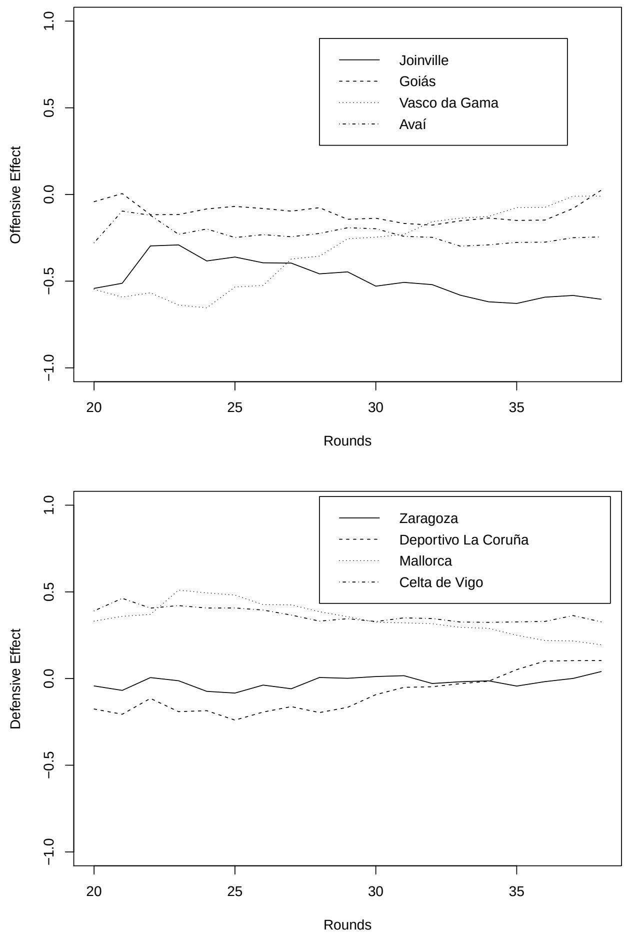

Figure E.2: *Attack and defense effect of the worst four teams in rounds 20–38.*

and worst four teams of the BFL in the 20–38 rounds. Corinthians and Atlético-MG have the highest attack effect. Corinthians also have the best defense effect. After 32-round Corinthians is the team with the best attack and defense effect. The four worst teams of BFL have an attack effect, meaning a low expected number of goals and few amount of victories that regulates these teams to the second division.

References

Baio G and Blangiardo M (2010). Bayesian hierarchical model for the prediction of football result, *Journal of Applied Statistics*, **37**, 253–264.

Bastos LS and da Rosa JMC (2013). Predicting probabilities for the 2010 FIFA world cup games using a Poisson-Gamma model, *Journal of Applied Statistics*, **40**, 1533–1544.

Brillinger DR (2008). Modelling game outcomes of the Brazilian 2006 series a championship as ordinal-valued, *Brazilian Journal of Probability Statistics*, **22**, 89–104.

Chib S and Greenberg E (1995). Understanding the Metropolis-Hastings algorithm, *The American Statistician*, **49**, 327–335.

Dixon MJ and Coles SG (1997). Modelling association football scores and inefficiencies in the football betting market, *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, **46**, 265–280.

Dyte D and Clarke SR (2000). A ratings based Poisson model for World Cup soccer simulation, *Journal of the Operational Research Society*, **51**, 993–998.

Gelman A, Carlin JB, Stern HS, and Rubin DB (1995). *Bayesian Data Analysis*, Chapman and Hall, London.

Gelman A, Sturtz S, Ligges U, Gorjane G, and Kerman J (2006). *The R2WinBUGS Package Manual Version 2.0-4*, Statistic Department Faculty, New York.

Gilks WR, Richardson S, and Spiegelhalter DJ (1996). *Markov Chain Monte Carlo in Practice*, Chapman and Hall, London.

Hastings WK (1970). Monte Carlo sampling methods using Markov Chains and their applications, *Biometrika*, **57**, 97–109.

Karlis D and Ntzoufras I (2003). Analysis of sports data by using bivariate Poisson models, *Journal of the Royal Statistical Society: Series D (The Statistician)*, **52**, 381–393.

Karlis D and Ntzoufras I (2009). Bayesian modelling of football outcomes: using the Skellam's distribution for the goal difference, *IMA Journal of Management Mathematics*, **20**, 133–145.

Keller JB (1994). A characterization of the Poisson distribution and the probability of winning a game, *The American Statistician*, **48**, 294–298.

Knorr-Held L (2000). Dynamic rating of sports teams, *Journal of the Royal Statistical Society: Series D (The Statistician)*, **49**, 261–276.

Koopman SJ and Lit R (2015). A dynamic bivariate Poisson model for analysing and forecasting match results in the English Premier League, *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, **178**, 167–186.

Lee AJ (1997). Modeling scores in the Premier League: is Manchester United really the best?, *Chance*, **10**, 15–19.

Maher MJ (1982). Modeling association football scores, *Statistica Neerlandica*, **36**, 109–118.

R Development Core Team (2012). *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria.

Spiegelhalter DJ, Thomas A, Best NG, and Lunn D (2003). *WinBUGS User Manual (Version 1.4.1)*, MRC Biostatistics Unit, Cambridge, UK.

Suzuki AK, Salasar LEB, Leite JG, and Louzada-Neto F (2010). A Bayesian approach for predicting match outcomes: the 2006 (Association) Football World Cup, *Journal of the Operational Research Society*, **61**, 1530–1539.

Volf P (2009). A random point process model for the score in sport matches, *IMA Journal of Management Mathematics*, **20**, 121–131.

Received May 5, 2016; Revised May 26, 2016; Accepted May 30, 2016