

Predicting Match Outcomes in the English Premier League: Which Will Be the Final Rank?

Francisco Louzada¹, Adriano K. Suzuki¹ and Luis E. B. Salasar²

¹*Universidade de São Paulo*

²*Universidade Federal de São Carlos*

Abstract: In any sport competition, there is a strong interest in knowing which team shall be the champion at the end of the championship. Besides this, the end result of a match, the chance of a team to be qualified for a specific tournament, the chance of being relegated, the best attack, the best defense, among others, are also subject of interest. In this paper we present a simple method with good predictive quality, easy implementation, low computational effort, which allows the calculation of all the interesting quantities above. Following Lee (1997), we estimate the average goals scored by each team by assuming that the number of goals scored by a team in a match follows a univariate Poisson distribution but we consider linear models that express the sum and the difference of goals scored in terms of five covariates: the goal average in a match, the home-team advantage, the team's offensive power, the opponent team's defensive power and a crisis indicator. The methodology is applied to the 2008-2009 English Premier League.

Key words: Linear Model, Premier Football League, Simulation, Poisson Distribution.

1. Introduction

Football is one of the most popular sports in the world. Played in several countries, it is a collective sport played by two teams whose purpose is to put the ball into the opposing team's goal (score a goal). The team which scores more goals wins the match. A draw occurs when the number of goals scored by the teams is the same. In several countries there are many football clubs competing for regional and national championships in several leagues. There are also intracontinental and intercontinental championships that can be played by teams which obtain qualification for.

This paper is directed to the English Premier League, which is one of the biggest and most valued national championships of football clubs. We particularly

concentrated on the 2008-09 Premier League season which was the seventeenth one since its establishment. During the course of a season twenty participating teams play in a single group; each team plays the others twice, one at their home stadium and the other at their opponents', for a total of 380 matches. The team that scores the most points at the end is declared the champion. If one or more teams finish with the same number of points, the teams are ranked according to the following order criteria: the goal difference followed by the goals scored. If a tie still persists, either in the championship, to qualify to other competitions or to relegation, a play-off match at a neutral venue decides the rank. The three best ranked teams qualify for the next (in our case the 2009-10) UEFA Champions League Group stage, the fourth, for the 2009-10 UEFA Champions League Play-off round, the fifth and the sixth, for the 2009-10 UEFA Europe League Play-off round and the seventh, for the 2009-10 UEFA Europe League Third qualifying round. The three worst ranked teams are relegated to the Football League Championship 2009-10.

In football, and also in any sports competition, there is a strong interest in knowing which team (in a collective sport) or which player (in an individual sport) shall be the champion at the end of the championship. Besides this, the end result of a match, the chance to qualify for a specific tournament, the chance of being relegated, the best attack, the best defense, among others, are also subject of interest.

Several papers are found in literature considering football score prediction applied to championship leagues such as the English Premier League (Lee 1997, Everson Goldsmith-Pinkham 2008, Karlis Ntzoufras 2009), the Norwegian Elite Division (Brillinger 2006) and the Brazilian Championship (Brillinger 2008). Lee (1997) considered a Poisson regression to model the number of goals from football team, where the average reflects the strength of the team, the quality of the opposition and the home advantage (if it is the home team). The independence between the goals scored by the two teams was assumed and his methodology was applied to the 1995-1996 English Premier League. Recently, Brillinger (2008) proposed to model directly the win, draw and loss probabilities. In that paper, Brillinger employed a trinomial model and applied it to the Brazilian 2006 Series A championship to obtain the estimate probability of any particular team to be champion, estimate the team's final points and to evaluate the chance of a team to be in the top four places. Karlis Ntzoufras (2009) applied the Skellam's distribution to model the goal difference between home and away teams. The authors argue that this approach does not rely neither on independence nor on the marginal Poisson distribution assumptions for the number of goals scored by the teams. A Bayesian analysis for predicting match outcomes for the English Premiere League (2006-2007 season) is carried out using a log-linear link function

and non-informative prior distributions for the model parameters.

In this paper, following Lee (1997), we estimate the average goals scored by each team by assuming that the number of goals scored by a team in a match follows a univariate Poisson distribution but we consider linear models that express the sum and the difference of goals scored in terms of five covariates: the goal average in a match, the home team advantage, the team's offensive power, the opponent team's defensive power and a crisis indicator. Generally, a football team may pass through a crisis when there is not a good relationship among players, between players and the coach or fans and the team, the coach criticizing player(s) and vice versa etc. This occurs most often when the team obtains successive negative results (successive losses and draws or bad performances).

The objective of this paper is to present a simple method with good predictive quality, easy implementation, low computational effort that allows the calculation of the interest probabilities: which team will be the champion, which ones will be relegated, which ones will qualify to another tournament, which team will be the best home team, which team will be the best away team (the team that scores the most points playing outside their hometown) etc. The model is applied to the 2008-2009 English Premier League. The Definetti measure (DeFinetti 1972) is used to quantify the model predictive quality.

To perform the forecasts we use directly from the Poisson model and the estimated means. We generate the score of the matches to be provide in order to estimate the simple matches predictions and also simulate several whole tournament to obtain the probability to be champion, to be relegated, to reach among the three best ranked teams etc.

The paper is outlined as follows. Section 2 presents the probabilistic model. Section 3 presents the results of our proposed modeling fitting applied to the 2008-09 UEFA Champions League. In this section we also quantify our modelling predictive quality by considering the Definetti measure (DeFinetti 1972). Section 4 presents the results of a simulation study performed for estimating some probabilities of interest such as, single match, champion, classification for the 2009-10 UEFA Champions League group phase and relegation. In Section 5 final considerations about the results and further work concludes the paper.

2. Probabilistic Model and Estimation

In this section we present the probabilistic model and estimation procedure. To illustrate we use as an example a fictitious tournament in which played the teams: Arsenal, Chelsea, Liverpool and Manchester United, whereas Liverpool is in crisis during the period in which the tournament is being played. In this tournament we will assume that occurred the following results

Liverpool 0×1 Arsenal, in Anfield Stadium;

Manchester United 2×1 Chelsea, in Old Trafford Stadium;

Chelsea 2×1 Liverpool, in Reebok Stadium;

Arsenal 0×2 Manchester United, in Emirates Stadium;

Chelsea 1×2 Arsenal, in Stamford Bridge Stadium;

Liverpool 0×3 Manchester United, in Reebok Stadium.

The final match of this tournament will be between Manchester United and Arsenal (the top two ranked teams) in Old Trafford Stadium. Based on the performance of teams in the tournament, the idea is to try to predict the possible chance of teams win, draw or lose the final match.

Thus, for a given match in a competition, consider X and Y the number of goals scored by the home and away teams. Henceforth, we shall assume X and Y independent Poisson distributed random variables with means λ_X and λ_Y , respectively.

From the linear properties of expectation, it follows that

$$\lambda_X = E[X] = \frac{E[X + Y] + E[X - Y]}{2} \quad (1)$$

and

$$\lambda_Y = E[Y] = \frac{E[X + Y] - E[X - Y]}{2}, \quad (2)$$

that is, λ_X and λ_Y can be expressed as functions of the means of the random variables $X + Y$ and $X - Y$. Observe that if $\text{Var}[X]$ and $\text{Var}[Y]$ are approximately equal, then the random variables $X + Y$ and $X - Y$ are uncorrelated, though X and Y are not necessarily uncorrelated, which can be seen as an advantage of modeling $X + Y$ and $X - Y$ instead of X and Y .

On the other hand, $E[X - Y]$ and $E[X + Y]$ are estimated by linear models given by

$$(X + Y)_i = S_i\alpha + \varepsilon_{ai} \quad (3)$$

and

$$(X - Y)_i = T_i\beta + \varepsilon_{bi}, \quad (4)$$

for $i = 1, 2, 3, \dots, n$, where n is the number of matches in the dataset; ε_{ai} and ε_{bi} are independent errors with mean equal to 0.

In linear model (3), $(X + Y)_i$ is the total of goals scored (by both teams) in the i -th match; the vector α is composed by $N + 2$ parameters (N parameters associated with each one of the teams listed in the dataset; one parameter associated with the home advantage and one parameter associated if one of the teams (or both) is in crisis).

The row vector S_i has $N + 2$ element, where N is associated to the status of each team over the match in question, one component that indicates the home advantage and one component that indicates if one of the teams (or both) is in crisis.

The status of a team is a variable incidence that takes the value 1 if it participates in the i -th match or 0 if not participating.

Assigning a common value for both teams involved in the match is due to the fact that the value of $(X + Y)_i$ not depend on identifying which team is X or Y . For example, the results 2×1 , 3×0 , 1×2 and 0×3 mean equally the occurrence of three goals ($(X + Y)_i = 3$).

The component relating to home advantage is also an incidence variable that can assume the values 1 if the match was played at the stadium of one of the teams or 0 it was in a neutral stadium. The crisis component is also an incidence variable that can assume the values 1 if one of the teams (or both) is in crisis or 0, otherwise.

For example the proposed application, in the first match of the tournament we have that $(X + Y)_1$ is equal to $0 + 1 = 1$ and the vector α given by $[\alpha_{Arse} \ \alpha_{Chel} \ \alpha_{Liver} \ \alpha_{Manc} \ \alpha_{Home} \ \alpha_{Crisis}]^t$, the row vector S_1 becomes equal to $[1 \ 0 \ 1 \ 0 \ 1 \ 1]$.

Considering jointly all tournament matches, $X + Y$ becomes the vector of goals total and S the matrix $n \times (N + 2)$ of status, local and crisis. For tournament matches, the model $(X + Y)_i = S_i\alpha + \varepsilon_{ai}$ is given by

$$\underbrace{\begin{bmatrix} 1 \\ 3 \\ 3 \\ 2 \\ 3 \\ 3 \end{bmatrix}}_{X + Y} = \underbrace{\begin{bmatrix} 1 & 0 & 1 & 0 & 1 & 1 \\ 0 & 1 & 0 & 1 & 1 & 0 \\ 0 & 1 & 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 & 1 & 0 \\ 1 & 1 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 1 & 0 & 1 \end{bmatrix}}_S \cdot \underbrace{\begin{bmatrix} \alpha_{Arse} \\ \alpha_{Chel} \\ \alpha_{Liver} \\ \alpha_{Manc} \\ \alpha_{Home} \\ \alpha_{Crisis} \end{bmatrix}}_\alpha + \varepsilon_a$$

for $\varepsilon_a = (\varepsilon_{a1}, \varepsilon_{a2}, \dots, \varepsilon_{an})^t$.

In a linear model (4), $(X - Y)_i$ is the difference of goals scored by teams X and Y in the i -th match; the vector β is composed by $N + 2$ parameters (N parameters associated with each one of the teams listed in the dataset; one parameter associated with the home advantage and one parameter associated if one of the teams (or both) is in crisis). From the practical point of view, in football, some teams play better as home team, others as an away team. In addition, a team that is going through a crisis has the pressure to achieve a positive result, which may impact on their performance in the match. Note that one team is in crisis in the i -th match but after some positive results or a major victory against a rival team can move out from this crisis status. The same applies if the opposite occurs. Motivated for these reasons we added these covariates in the model, relating them to the number of goals scored by the teams X and Y .

The row vector T_i has $N + 2$ element, where N is associated to the status of each team over the match in question, one component that indicates the home advantage and one component that indicates if one of the teams (or both) is in crisis.

In this model, if the team participates in the i -th match the status is a variable incidence that takes the value 1 (for team X) and -1 (for team Y). If not participating, the status is equal to 0. It is necessary to distinguish between teams X and Y because the value of $(X - Y)_i$ depend directly identifying which team is the X and Y . For example, the results 3×2 and 2×3 have meanings completely different because $(X - Y)_i = 1$ and $(X - Y)_i = -1$, respectively. The component relating to home advantage is also an incidence variable that can assume the values 1 if the match was played at the stadium of one of the teams or 0 it was in a neutral stadium. The crisis component is also an incidence variable that can assume the values 1 if one of the teams (or both) is in crisis or 0, otherwise.

For example the proposed application, in the fourth match of the tournament we have that $(X - Y)_4$ is equal to $0 - 2 = -2$ and the vector β given by

$[\beta_{Arse} \ \beta_{Chel} \ \beta_{Liver} \ \beta_{Manc} \ \beta_{Home} \ \beta_{Crisis}]^t$, the row vector T_4 becomes equal to $T_4 = [1 \ 0 \ 0 \ -1 \ 1 \ 0]$.

Considering jointly all tournament matches, $X - Y$ becomes the vector of goals difference and T the matrix $n \times (N + 2)$ of status, local and crisis. For tournament matches, the model $(X - Y)_i = T_i\beta + \varepsilon_{bi}$ is given by

$$\underbrace{\begin{bmatrix} -1 \\ 1 \\ 1 \\ -2 \\ -1 \\ -3 \end{bmatrix}}_{X - Y} = \underbrace{\begin{bmatrix} -1 & 0 & 1 & 0 & 1 & 1 \\ 0 & -1 & 0 & 1 & 1 & 0 \\ 0 & 1 & -1 & 0 & 0 & 1 \\ 1 & 0 & 0 & -1 & 1 & 0 \\ -1 & 1 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & -1 & 0 & 1 \end{bmatrix}}_T \cdot \underbrace{\begin{bmatrix} \beta_{Arse} \\ \beta_{Chel} \\ \beta_{Liver} \\ \beta_{Manc} \\ \beta_{Home} \\ \beta_{Crisis} \end{bmatrix}}_{\beta} + \varepsilon_b$$

for $\varepsilon_b = (\varepsilon_{b1}, \varepsilon_{b2}, \dots, \varepsilon_{bn})^t$.

Thus, in example, the estimators of λ_1 and λ_2 for the final match between Manchester United and Arsenal, built in can be calculated from $\widehat{E}[X + Y] = S_7\widehat{\alpha}$ and $\widehat{E}[X - Y] = T_7\widehat{\beta}$.

So, for the final match between Manchester United and Arsenal held in Old Trafford Stadium, the vectors S_7 and B_7 are given by $S_7 = [1 \ 0 \ 0 \ 1 \ 1 \ 0]$ and $T_7 = [-1 \ 0 \ 0 \ 1 \ 1 \ 0]$, where we obtain the point forecasts $\widehat{E}[X + Y] = S_7\widehat{\alpha} = \widehat{\alpha}_{Man} + \widehat{\alpha}_{Arse} + \widehat{\alpha}_{Home}$ and $\widehat{E}[X - Y] = T_7\widehat{\beta} = \widehat{\beta}_{Man} - \widehat{\beta}_{Arse} + \widehat{\beta}_{Home}$, where X and Y are, respectively, the number of scored goals by Manchester United and Arsenal held in Old Trafford Stadium.

Assuming equal weights, from the Moore-Penrose generalized inverse matrix (Venables Ripley, 1999) we obtain the desired estimatives $\widehat{\alpha}$ and $\widehat{\beta}$ given by

$$\widehat{\alpha} = (S'S)^{-1}S'(X + Y) = [1.750 \ 2.500 \ 0.375 \ 2.000 \ -1.500 \ 0.375]^t$$

and

$$\widehat{\beta} = (T'T)^{-1}T'(X - Y) = [-0.097 \ -0.282 \ -1.153 \ 1.532 \ 1.532 \ 1.532]^t.$$

Soon,

$$\widehat{E}[X + Y] = \widehat{\alpha}_{Man} + \widehat{\alpha}_{Arse} + \widehat{\alpha}_{Home} = 2.000 + 1.750 - 1.500 = 2.250$$

and

$$\widehat{E}[X - Y] = \widehat{\beta}_{Man} - \widehat{\beta}_{Arse} + \widehat{\beta}_{Home} = 1.532 + 0.097 - 0.518 = 1.11.$$

Thus, we obtain the estimates

$$\widehat{\lambda}_{Manc} = \frac{\widehat{E}[X + Y] + \widehat{E}[X - Y]}{2} = \frac{2.25 + 1.11}{2} = 1.68 \tag{5}$$

and

$$\hat{\lambda}_{Arse} = \frac{\hat{E}[X + Y] - \hat{E}[X - Y]}{2} = \frac{2.25 - 1.11}{2} = 0.57. \quad (6)$$

Then, the “least squares score” for the final match would be (1.68×0.57) .

Note that we can encounter a problem when $\hat{E}[X - Y] > \hat{E}[X + Y]$. When this occurs, we obtain $\hat{\lambda}_Y < 0$ in expression (2). To overcome this problem, we use the “walk” to the set of valid estimative. If we equate $\hat{\lambda}_Y$ to the nearest value of valid estimative, i.e, $\hat{\lambda}_Y = 0$, we would be assuming that the team Y has no chance of winning the match, what in football is not feasible, since even a team with technical quality much lower than the opponent, we can observe the possibility of occurrence of this team win (unexpected result). Thus, in the few situations where this problem occurred, we equate the rate of team Y to a small value, i.e., $\hat{\lambda}_Y = 0.25$.

2.1 Deriving the probabilities

For a given match played by teams X and Y , we calculate the probabilities of win (P_W), draw (P_D) and loss (P_L) of team A from the predictive distributions, using the following equations

$$P_W = P(X > Y) = \sum_{i=1}^{\infty} \sum_{j=0}^{i-1} P(X = i)P(Y = j), \quad (7)$$

$$P_D = P(X = Y) = \sum_{i=0}^{\infty} P(X = i)P(Y = i), \quad (8)$$

and

$$P_L = P(X < Y) = \sum_{j=1}^{\infty} \sum_{i=0}^{j-1} P(X = i)P(Y = j). \quad (9)$$

From the values of $\hat{\lambda}_{Manc}$ and $\hat{\lambda}_{Arse}$ obtained in (5) and (6) we can calculate the probability of Manchester win, draw and Arsenal win through the expressions (7), (8) and (9), respectively. Thus, we obtain the following probabilities

$$P[\text{Manchester win}] = 0.6449,$$

$P[\text{draw}] = 0.2332$ and

$P[\text{Arsenal win}] = 0.1219$.

2.2 Algorithm for the simulation

Suppose the tournament is composed by N rounds. For each round r , $r = N/2, \dots, N$, we obtained the final team classification, i.e., number of points, number of victories, number of draws, number of defeat, number of goals scored, number of goals conceded and goal differences. The forecast for the final classification was performed using a simulation based on Poisson model involving the following steps

a) Fix n the number of championships to be simulated and r the number of round played. Do $c = 1$ (the counter);

b) If $c < n$ use the $(r - 1) * 10$ observed matches to estimate the home and away teams goal rates;

c) For each one of $M = [N - (r - 1)] * 10$ matches to be played, simulate the number of goals scored using the Poisson distribution with estimated rates obtained in step **(b)**. Do $c = c + 1$ and return to step **(b)**.

To assemble the final league tables, for each M matches predicted, see if there was victory of team X ($X > Y$), draw ($X = Y$) or victory of team Y ($X < Y$). Give 3 points for the winning team and 1 point for both teams if there was a draw. From the current league table, update with the simulated results for each of the n simulated championships. By the final league tables, we can calculate, for example, the chance of a particular team to be champion and to be relegation as follows

$P[\text{Team to be champion}] = \#(\text{team finished in the first place in the final league table})/n$,

$P[\text{Team to be relegation}] = \#(\text{team finished in the last three placed in the final league table})/n$,

where $\#$ refers to the number of times obtained in the simulation.

3. Data Analysis

In this section we present the results obtained by applying your methodology

to the English Premier League, particularly to the 2008-09 Premier League season. We focus on the single match predictions as well as on the predictions for the whole Tournament.

3.1 General Data Structure and Assumption

We used as our data set the outcomes of the 180 first matches (18 rounds played) of the 2008-09 Premier League season to perform the predictions of the following 200 matches (from 19th to 38th rounds). The first 18 rounds were chosen as the training set since only after 18 rounds we have observed one match of each team against to the other opponents. The team crisis indicator, stated for each team in each round, was based on the media information about the team.

In order to apply our methodology, we argue that the number of goals scored by each team in a match follows a univariate Poisson distribution, and the assumption of independence is presumed. We however confirm such assumption through a naive Chi-square (χ^2) test.

Considering the number of goals scored by the home team we observed $\chi_{obs}^2 = 2.1313$ on 5 degree of freedom with the critical value equals to $\chi_c^2 = 15.0863$ at 1% of significance. Considering the number of goals scored by the away team we observed $\chi_{obs}^2 = 10.3086$ on 4 degree of freedom with the critical value given by $\chi_c^2 = 13.2767$ at 1% of significance. Therefore, as in both cases we have smaller observed values than the critical ones there is insufficient evidence to reject the hypothesis that the goals follow a Poisson distribution. The test which shows that there is no evidence against the assumption of independence was performed through a cross-tabulation of the home and away number of goals scored of all 380 matches in the 2008-2009 English Premier League, with an observed value equals to $\chi_{obs}^2 = 16.9325$ on 16 degree of freedom, which is smaller than the critical one $\chi_c^2 = 31.9999$ at 1% of significance.

3.2 Quality of the Predictions

As pointed out in Section 1, the Definetti measure (DeFinetti 1972) was used to quantify our modeling predictive quality. Consider the set of all possible forecasts given by the simplex set $S = \{(P_W, P_D, P_L) \in [0, 1]^3 : P_W + P_D + P_L = 1\}$, where P_W denotes the win probability, P_D denotes the draw probability and P_L denotes the loss probability. The vertices $(1, 0, 0)$, $(0, 1, 0)$ and $(0, 0, 1)$ of S represent the outcomes win, draw and loss, respectively. Thus, following (DeFinetti 1972), we calculate the DeFinetti distance, which is the Euclidean distance, between the point corresponding to the outcome and that one corresponding to the prediction. For instance, if a prediction is given by $(0.45, 0.20, 0.35)$ and the outcome is a victory $(1, 0, 0)$, then the DeFinetti dis-

tance is given by $(0.45 - 1)^2 + (0.20 - 0)^2 + (0.35 - 0)^2 = 0.465$. Also, we can associate the average of their DeFinetti distances to a set of predictions, known as the DeFinetti measure.

Before each one of the 20 remaining rounds (19th to 38th rounds) we calculated, the win, draw and loss probabilities (see Subsection 2.1) for all matches and the Defineti measure (DeFinetti 1972) associated with these predictions.

For each round, the associated Defineti measure obtained was 0.479, 0.510, 0.667, 0.508, 0.457, 0.700, 0.502, 0.650, 0.629, 0.498, 0.680, 0.645, 0.432, 0.545, 0.568, 0.499, 0.312, 0.575, 0.529 and 0.332. We can observe that in almost all rounds the Defineti measures were less than $2/3$ equivalent to an equiprobable predictor (which assigns equal probability for all outcomes $P_W = P_D = P_L = 1/3$). Moreover, in the modeling we consider the two covariates home and crisis status covariates, relating them to the number of goals scored by the teams X and Y. To check if these covariates impact the modeling, we fit the model again but removing one or both covariates of the model and calculate the probabilities of victory, draw and loss and also the corresponding Defineti measure. A better prediction, which is the main interest here, was obtained by considering the full model with the two covariates.

3.3 Single Match Prediction

In this section, we present the forecasts for all the matches of the 35th round which are shown in Table 1. If we calculate the percentage of correct forecasts, a forecast (P_W, P_D, P_L) shall be considered correct if the outcome with the greatest probability coincides with the observed outcomes, then our model scored 9 results with a Defineti measure equals to 0.3119.

Table 1: Forecasts for single matches of the 35th round.

Home Team	Observed Score	Away Team	W	D	L	Correct
Chelsea	3 - 1	Fulham	0.6910	0.2700	0.039	Yes
Tottenham Hotspurs	1 - 0	West Bromwich Albion	0.6263	0.2156	0.1581	Yes
Wigan Athletic	0 - 0	Bolton Wanderers	0.4773	0.2732	0.2495	No
Portsmouth	0 - 3	Arsenal	0.2038	0.2343	0.5619	Yes
Stoke City	0 - 1	West Ham United	0.3316	0.2975	0.3709	Yes
Middlesbrough	0 - 2	Manchester United	0.0591	0.1904	0.7505	Yes
Manchester City	3 - 1	Blackburn Rovers	0.6122	0.2051	0.1827	Yes
Liverpool	3 - 0	Newcastle United	0.7863	0.1518	0.0619	Yes
Sunderland	0 - 2	Everton	0.2479	0.2836	0.4685	Yes
Aston Villa	1 - 0	Hull City	0.6238	0.2037	0.1725	Yes

3.4 Predictions for the whole Tournament

In this section, we present the final prediction of the classification on the

tournament. The predictions of all the matches before each round were made. The champion of the 2008-2009 English Premier League was Manchester United that obtained the most overall wins (28 times) and had the fewest goals conceded (24 goals). It was the best as the home team (16 victories and 43 goals scored) and the one with the fewest goals conceded as the away team (11).

We applied our modelling in order to calculate the probabilities of all matches, for instance, from 30th round (90 matches). The predicted and observed (in brackets) final classification are presented in Table 2. We can observe that our modeling approaches very well the final classification of the tournament.

Table 2: Final classification table assembled of all matches from 30 rounds (Final tournament classification).

Team	Played	Won	Drawn	Lost	Goals for	Goals against	Goal difference	Points
Manchester United	38	27 (28)	6 (6)	5 (4)	69 (68)	28 (24)	41 (44)	87 (90)
Chelsea	38	24 (25)	8 (8)	6 (5)	71 (68)	37 (24)	34 (44)	80 (83)
Liverpool	38	23 (25)	11 (11)	4 (2)	63 (77)	37 (27)	26 (50)	80 (86)
Arsenal	38	19 (20)	12 (12)	7 (6)	60 (68)	40 (37)	20 (31)	69 (72)
Everton	38	19 (17)	12 (12)	7 (9)	56 (55)	32 (37)	24 (18)	69 (63)
Aston Villa	38	18 (17)	8 (11)	12 (10)	55 (54)	56 (48)	-1 (6)	62 (62)
Manchester City	38	16 (15)	5 (5)	17 (18)	62 (58)	53 (50)	9 (8)	53 (50)
Fulham	38	13 (14)	12 (11)	13 (13)	37 (39)	52 (34)	-15 (5)	51 (53)
West Ham United	38	13 (14)	9 (9)	16 (15)	41 (42)	45 (45)	-4 (-3)	48 (51)
Tottenham Hotspur	38	13 (14)	8 (9)	17 (15)	46 (45)	56 (45)	-10 (0)	47 (51)
Sunderland	38	13 (9)	8 (9)	17 (20)	41 (34)	39 (54)	2 (-20)	47 (36)
Blackburn Rovers	38	12 (10)	10 (11)	16 (17)	48 (40)	57 (60)	-9 (-20)	46 (41)
Wigan Athletic	38	12 (12)	9 (9)	17 (17)	36 (34)	36 (45)	0 (-11)	45 (45)
Bolton Wanderers	38	13 (11)	4 (8)	21 (19)	44 (41)	60 (53)	-16 (-12)	43 (41)
Hull City	38	11 (8)	10 (11)	17 (19)	48 (39)	44 (64)	4 (-25)	43 (35)
Stoke City	38	11 (12)	8 (9)	19 (17)	39 (38)	64 (55)	-25 (-17)	41 (45)
Newcastle United	38	8 (7)	13 (13)	17 (18)	42 (40)	46 (59)	-4 (-19)	37 (34)
Portsmouth	38	9 (10)	9 (11)	20 (17)	38 (38)	46 (57)	-8 (-19)	36 (41)
Middlesbrough	38	7 (7)	12 (11)	19 (20)	29 (28)	53 (57)	-24 (-29)	33 (32)
West Bromwich Albion	38	9 (8)	6 (8)	23 (22)	39 (36)	55 (67)	-16 (-31)	33 (32)

4. Competition Simulation

In this section we present the others probabilities of interest: champion, classification to the 2009-10 UEFA Champions League group phase and relegation.

4.1 General Specification

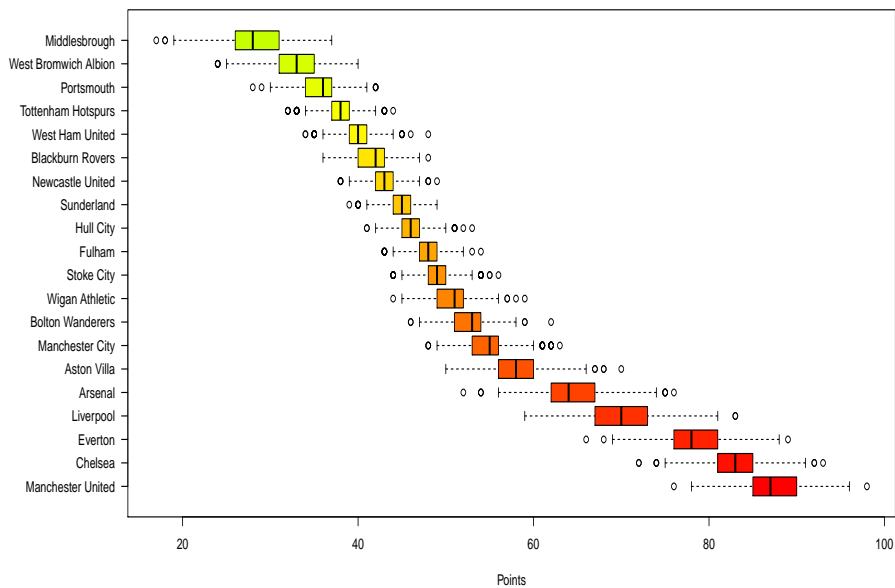
We consider a 1,000 tournament replications. A tournament replica was obtained by considering the simulation procedure briefly described above. Then, calculate the percentage of tournament replica wins for each team, the percentage of tournament replica to be qualified for the 2009-10 UEFA Champions League Group stage (to be among the three best ranked teams), to be relegated to the Football League Championship 2009-10 (to be among the three worst ranked

teams) and to end up in the last place.

4.2 Overall Results

In this subsection, based on our 1,000 tournament replications, several types of interesting information can be obtained, such as: how many times a team was the champion, how many times a team finished in the first three positions, the variability of the number of points, goals scored, goals taken, the number of victories, losses, draws etc.

All the results are presented in terms of average. Initially, based on the observed data before the 20th round, Figure 1 displays the box-plots of the 1,000 predicted numbers of points for each team at the final of the tournament. We observe that based on the available data, the fitted model indicates Manchester United as favorite team for winner of the tournament, followed by Chelsea, Everton and Liverpool. In fact, Manchester United won the tournament and Liverpool finished at second place, indicating an improvement in the performance of the Liverpool team. Also, Tables 6 and 7 at the Appendix, displays the probabilities of each one of the 20 positions that each team reached at the end of the championship by considering the observed data before the 25th and 30th rounds, respectively.



Box-plot of the number of points obtained by each team before 20th round.

4.3 Some Specific Results

In this section, before each of the 20 remaining rounds (19th to 38th rounds), a simulation of 50 times 1,000 whole tournaments was performed to obtain the mean and the standard deviation of the probabilities. The tables below show the probabilities of tournament wins (Table 3), to reach among the top three places (Table 4) and to be relegated (Table 5), for each of the 20 remaining rounds (19th to 38th rounds) for each team with the highest probabilities.

We observed in Table 3 that for all rounds the champion Manchester United is the team which most finished in the first place in our simulation.

Table 3: Simulation results for the five teams with the the highest percentages of tournament wins.

Round	Manchester United		Liverpool		Chelsea		Arsenal		Aston Villa	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
19	0.308	0.012	0.073	0.007	0.616	0.014	0.001	0.001	0.002	0.001
20	0.300	0.018	0.120	0.009	0.578	0.018	0.000	0.001	0.002	0.001
21	0.328	0.014	0.272	0.015	0.398	0.013	0.000	0.000	0.002	0.002
22	0.695	0.018	0.143	0.012	0.159	0.013	0.000	0.001	0.002	0.002
23	0.755	0.014	0.080	0.009	0.160	0.012	0.001	0.001	0.004	0.002
24	0.856	0.011	0.032	0.006	0.109	0.009	0.000	0.000	0.003	0.002
25	0.900	0.012	0.066	0.009	0.033	0.006	0.000	0.000	0.001	0.001
26	0.930	0.008	0.061	0.008	0.008	0.003	0.000	0.000	0.001	0.001
27	0.943	0.007	0.037	0.006	0.019	0.004	0.000	0.000	0.000	0.001
28	0.977	0.005	0.005	0.003	0.017	0.005	0.000	0.000	0.000	0.000
29	0.979	0.004	0.005	0.002	0.016	0.004	0.000	0.000	0.000	0.000
30	0.879	0.012	0.051	0.007	0.070	0.009	0.000	0.000	0.000	0.000
31	0.737	0.015	0.224	0.013	0.040	0.006	0.000	0.000	0.000	0.000
32	0.742	0.016	0.232	0.015	0.026	0.005	0.000	0.000	0.000	0.000
33	0.779	0.013	0.208	0.013	0.012	0.004	0.000	0.000	0.000	0.000
34	0.926	0.007	0.072	0.007	0.001	0.001	0.000	0.000	0.000	0.000
35	0.946	0.007	0.053	0.007	0.001	0.001	0.000	0.000	0.000	0.000
36	0.963	0.006	0.037	0.006	0.000	0.000	0.000	0.000	0.000	0.000
37	0.988	0.004	0.012	0.004	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	0.000	0.000	0.000

To qualify for the 2009-10 UEFA Champions League Group stage, the teams need to finish among the three best ranked teams. The probabilities of the six teams that had some probability of reaching the top three places are presented in Table 4. In Table 4, the teams that qualified (Manchester United, Liverpool and Chelsea) had a higher probabilities of finishing among the three best ranked teams.

Another probability of interest is the one of the teams that will be relegated. In a round-robin tournament, there is an extensive dispute, both to be champion or to qualify for any tournament, but also not to be relegated. The teams are relegated when they finish among the three worst ranked teams. The probabilities

Table 4: Simulation results for the six teams that had some percentage in reaching the top three places.

Round	Manchester United		Liverpool		Chelsea		Arsenal		Aston Villa		Everton	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
19	0.976	0.005	0.844	0.011	0.993	0.003	0.055	0.006	0.133	0.011	0.000	0.000
20	0.977	0.005	0.917	0.009	0.992	0.003	0.033	0.006	0.080	0.008	0.000	0.000
21	0.969	0.005	0.958	0.006	0.979	0.004	0.023	0.005	0.071	0.008	0.000	0.001
22	0.995	0.002	0.923	0.009	0.941	0.009	0.041	0.007	0.099	0.009	0.001	0.001
23	0.994	0.002	0.844	0.010	0.926	0.009	0.072	0.007	0.163	0.013	0.001	0.001
24	0.998	0.001	0.784	0.016	0.927	0.008	0.058	0.009	0.232	0.016	0.000	0.001
25	0.999	0.001	0.936	0.008	0.887	0.009	0.031	0.005	0.146	0.011	0.000	0.001
26	1.000	0.001	0.946	0.007	0.759	0.014	0.023	0.005	0.272	0.013	0.002	0.001
27	1.000	0.000	0.934	0.009	0.909	0.010	0.008	0.003	0.149	0.012	0.000	0.001
28	1.000	0.000	0.882	0.012	0.966	0.005	0.009	0.003	0.141	0.013	0.002	0.001
29	1.000	0.000	0.946	0.008	0.989	0.003	0.011	0.003	0.053	0.008	0.000	0.001
30	1.000	0.000	0.991	0.003	0.997	0.002	0.007	0.003	0.004	0.002	0.000	0.000
31	1.000	0.000	0.998	0.002	0.987	0.003	0.014	0.004	0.001	0.001	0.000	0.000
32	1.000	0.000	0.999	0.001	0.988	0.003	0.013	0.003	0.000	0.000	0.000	0.000
33	1.000	0.000	0.999	0.001	0.976	0.005	0.025	0.005	0.000	0.000	0.000	0.000
34	1.000	0.000	1.000	0.001	0.967	0.005	0.034	0.005	0.000	0.000	0.000	0.000
35	1.000	0.000	1.000	0.000	0.987	0.003	0.013	0.004	0.000	0.000	0.000	0.000
36	1.000	0.000	1.000	0.000	0.992	0.003	0.008	0.003	0.000	0.000	0.000	0.000
37	1.000	0.000	1.000	0.000	1.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
38	1.000	0.000	1.000	0.000	1.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000

of the five teams that had more probability in finishing in the last three places are presented in Table 5. The relegated teams were West Bromwich Albion, Middlesbrough and Newcastle United. In Table 5, it is observed that West Bromwich Albion and Middlesbrough (the two last ranked teams) had the highest percentage to be relegated in all rounds. As for the other teams, there was a variation in these values.

5. Final Remarks

In this paper we proposed a simple method with good predictive quality, easy implementation and low computational effort for predicting match outcomes. It presents some probabilities of interest that can be obtained such as, simple match, champion, classification for the 2009-10 UEFA Champions League group phase and relegation. However, other results can be obtained, such as, the chance of each team to finish as the last ranked team, will any team qualify for the 2009-10 UEFA European League? Which team will have the best defense (take less goals)? Which team will have the best attack (score more goals)? Which team will win more matches? What will the best home team be? What will the best away team be? etc. The methodology is implemented in the software R and interested readers can have the code by emailing the authors.

Although our modelling was applied to the 2008-2009 English Premier League, in principle, it is flexible and can be easily adapted to other different tournaments.

After the tournament we can observe a higher occurrence of the home team's

Table 5: Simulation results of the five teams that had more percentage in finishing in the last three places.

Round	West Bromwich Albion		Sunderland		Newcastle United		Middlesbrough		Hull City	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
19	0.943	0.008	0.067	0.007	0.023	0.004	0.402	0.018	0.007	0.003
20	0.930	0.008	0.087	0.010	0.038	0.006	0.488	0.014	0.018	0.004
21	0.780	0.012	0.166	0.010	0.078	0.009	0.425	0.014	0.022	0.005
22	0.832	0.012	0.205	0.011	0.101	0.009	0.497	0.014	0.055	0.007
23	0.620	0.015	0.269	0.016	0.228	0.013	0.739	0.013	0.078	0.008
24	0.724	0.012	0.125	0.011	0.265	0.013	0.740	0.014	0.119	0.010
25	0.742	0.012	0.128	0.011	0.287	0.013	0.793	0.013	0.141	0.010
26	0.816	0.012	0.044	0.006	0.154	0.011	0.824	0.011	0.085	0.009
27	0.893	0.011	0.028	0.005	0.145	0.010	0.817	0.012	0.141	0.011
28	0.939	0.007	0.038	0.006	0.275	0.016	0.580	0.019	0.284	0.014
29	0.971	0.004	0.059	0.007	0.337	0.014	0.799	0.012	0.120	0.010
30	0.958	0.007	0.079	0.008	0.328	0.016	0.790	0.013	0.095	0.010
31	0.973	0.005	0.096	0.010	0.446	0.015	0.880	0.009	0.112	0.009
32	0.993	0.003	0.166	0.012	0.557	0.015	0.957	0.007	0.125	0.010
33	0.993	0.002	0.175	0.014	0.550	0.015	0.874	0.011	0.214	0.013
34	0.997	0.002	0.047	0.007	0.671	0.015	0.854	0.011	0.287	0.015
35	0.988	0.003	0.109	0.010	0.709	0.012	0.879	0.010	0.298	0.014
36	0.996	0.002	0.122	0.012	0.675	0.013	0.886	0.010	0.310	0.017
37	0.981	0.004	0.060	0.008	0.309	0.014	0.974	0.005	0.674	0.013
38	1.000	0.000	0.018	0.004	0.658	0.015	0.989	0.003	0.335	0.015

victories (in the 2009-10 UEFA Champions League case, 173 victories for the home teams, the 110 victories for the away teams and 97 draws). However, before each round, we do not know the match outcomes, then assign an equal chance to the three possible outcomes (win, draw, loss) seems a reasonable strategy that is independent of subjectivity. However, other possibilities should be considered further. For instance, if one knows in advance the chances of win, draw and loss, but note that in this case, subjectivity is taken into account.

For simplicity we adopt here a linear model relating the goals scored by the teams and the covariates. However, a polynomial could be adopted. This can be seen as a direct generalization of our modeling and may be investigated further. Moreover, intuitively, if several matches are played within the same conditions: home field advantage, crisis, judge, atmospheric condition, the presence of dependence may exist. However, such situation was not considered here but should be investigated further in the context of our modeling. Particularly, we may assume a bivariate distribution for (X, Y) and consequently it will be possible to check for the presence of dependence between them.

Acknowledgements

The authors are grateful to anonymous referees and the Editor for very useful comments and suggestions. The research of Francisco Louzada is supported by

the Brazilian organization CNPq.

Appendix

Reference

- Brillinger, D. R. (2006). Modelling some Norwegian soccer data. *Advances in Statistical Modelling and Inference*, (Ed. V. J. Nair.) World Scientific., pages 3-20.
- Brillinger, D. R. (2008). Modeling game outcomes of the Brazilian 2006 series a championship as ordinalvalued. *Braz J Probab Stat*, **22(2)**, 89-104.
- DeFinetti, B. (1972). *Probability, Induction and Statistics*. John Wiley, London.
- Everson, P. Goldsmith-Pinkham, P. (2008). Composite Poisson models for goal scoring. *J Quant Anal Sports*, **4(2)**.
- Karlis, D. Ntzoufras, I. (2009). Bayesian modeling of football outcomes: Using the Skellam's distribution for the goal difference. *IMA J Manag Math*, **20**, 133-145.
- Lee, A. (1997). Modeling scores in the Premier League: Is Manchester United really the best? *Chance*, **10(1)**, 15-19.
- Venables, W. N. Ripley, B. D. (1999). *Modern Applied Statistics with S-PLUS*. Third Edition. Springer p.100.

Table 6: Mean, standard deviation of final league ranks, and probability of each ranking (in %)- 25th round.

Team	Summary		Probability of each ranking																				
	Mean	SD	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	
Manchester United	1.126	0.385	0.892	0.091	0.016	0.001	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Liverpool	2.413	0.742	0.072	0.517	0.345	0.058	0.008	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Chelsea	2.712	0.757	0.034	0.358	0.482	0.114	0.012	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Arsenal	4.808	0.718	0.000	0.003	0.031	0.243	0.614	0.100	0.006	0.002	0.001	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Everton	6.297	0.939	0.000	0.000	0.001	0.015	0.093	0.597	0.209	0.057	0.015	0.010	0.003	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Aston Villa	4.103	0.774	0.002	0.031	0.125	0.568	0.253	0.020	0.001	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Fulham	11.386	2.509	0.000	0.000	0.000	0.000	0.000	0.001	0.030	0.079	0.137	0.156	0.168	0.129	0.112	0.063	0.049	0.033	0.028	0.013	0.002	0.000	0.000
Tottenham Hotspur	15.021	2.856	0.000	0.000	0.000	0.000	0.000	0.001	0.010	0.021	0.036	0.061	0.076	0.092	0.113	0.120	0.123	0.132	0.094	0.086	0.034	0.000	0.000
West Ham United	8.848	2.006	0.000	0.000	0.000	0.000	0.003	0.061	0.203	0.247	0.178	0.134	0.075	0.049	0.023	0.008	0.006	0.009	0.003	0.000	0.001	0.000	0.000
Manchester City	9.916	2.336	0.000	0.000	0.000	0.000	0.000	0.025	0.086	0.187	0.210	0.170	0.114	0.071	0.049	0.032	0.030	0.013	0.008	0.003	0.001	0.000	0.000
Wigan Athletic	7.519	1.401	0.000	0.000	0.000	0.001	0.016	0.184	0.400	0.216	0.101	0.039	0.025	0.010	0.004	0.003	0.001	0.000	0.000	0.000	0.000	0.000	0.000
Stoke City	15.082	3.017	0.000	0.000	0.000	0.000	0.000	0.002	0.007	0.035	0.035	0.060	0.078	0.092	0.097	0.125	0.111	0.103	0.112	0.081	0.062	0.004	0.000
Bolton Wanderers	11.378	2.848	0.000	0.000	0.000	0.000	0.001	0.007	0.044	0.105	0.141	0.146	0.114	0.119	0.095	0.077	0.057	0.039	0.025	0.019	0.007	0.004	0.000
Portsmouth	14.154	2.918	0.000	0.000	0.000	0.000	0.000	0.001	0.003	0.018	0.037	0.054	0.092	0.100	0.103	0.130	0.123	0.106	0.097	0.061	0.049	0.026	0.000
Blackburn Rovers	14.973	2.799	0.000	0.000	0.000	0.000	0.000	0.001	0.008	0.017	0.043	0.059	0.073	0.108	0.117	0.114	0.130	0.113	0.113	0.073	0.031	0.000	0.000
Sunderland	13.868	2.959	0.000	0.000	0.000	0.000	0.000	0.001	0.006	0.028	0.049	0.058	0.083	0.109	0.116	0.131	0.101	0.101	0.088	0.079	0.037	0.013	0.000
Hull City	14.000	3.076	0.000	0.000	0.000	0.000	0.000	0.002	0.006	0.027	0.046	0.068	0.088	0.100	0.097	0.103	0.115	0.110	0.088	0.088	0.043	0.020	0.000
Newcastle United	15.543	2.948	0.000	0.000	0.000	0.000	0.000	0.002	0.009	0.011	0.046	0.045	0.065	0.082	0.083	0.094	0.128	0.118	0.154	0.101	0.062	0.000	0.000
Middlesbrough	18.529	1.751	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.004	0.006	0.012	0.018	0.033	0.045	0.105	0.130	0.275	0.372	0.000
West Bromwich Albion	18.324	2.094	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.001	0.005	0.009	0.015	0.015	0.025	0.032	0.052	0.093	0.134	0.244	0.375	0.000	0.000

Table 7: Mean, standard deviation of final league ranks, and probability of each ranking (in %) - 30th round.

Team	Summary		Probability of each ranking																				
	Mean	SD	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	
Manchester United	1.132	0.401	0.891	0.086	0.023	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Liverpool	2.493	0.607	0.053	0.407	0.534	0.006	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Chelsea	2.383	0.594	0.056	0.507	0.435	0.002	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Arsenal	4.581	0.684	0.000	0.000	0.005	0.515	0.375	0.104	0.001	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Everton	5.697	0.802	0.000	0.000	0.000	0.093	0.215	0.614	0.063	0.010	0.005	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Aston Villa	4.821	0.764	0.000	0.000	0.003	0.383	0.409	0.200	0.005	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Fulham	10.222	1.869	0.000	0.000	0.000	0.000	0.000	0.003	0.054	0.130	0.183	0.214	0.186	0.116	0.064	0.029	0.014	0.006	0.001	0.000	0.000	0.000	0.000
Tottenham Hotspur	11.200	2.216	0.000	0.000	0.000	0.000	0.000	0.001	0.037	0.062	0.114	0.184	0.188	0.165	0.109	0.065	0.033	0.023	0.010	0.006	0.003	0.000	0.000
West Ham United	8.464	1.542	0.000	0.000	0.000	0.000	0.030	0.274	0.293	0.189	0.113	0.052	0.028	0.013	0.007	0.000	0.001	0.000	0.000	0.000	0.000	0.000	0.000
Manchester City	9.238	1.834	0.000	0.000	0.000	0.000	0.000	0.162	0.219	0.225	0.155	0.108	0.064	0.029	0.013	0.009	0.002	0.001	0.000	0.000	0.000	0.000	0.000
Wigan Athletic	8.229	1.573	0.000	0.000	0.000	0.001	0.001	0.034	0.389	0.231	0.159	0.094	0.052	0.019	0.011	0.006	0.002	0.000	0.001	0.000	0.000	0.000	0.000
Stoke	15.633	2.422	0.000	0.000	0.000	0.000	0.000	0.001	0.001	0.005	0.021	0.027	0.063	0.079	0.117	0.136	0.128	0.161	0.159	0.082	0.020	0.000	0.000
Bolton Wanderers	12.800	2.410	0.000	0.000	0.000	0.000	0.000	0.001	0.009	0.027	0.056	0.075	0.120	0.171	0.163	0.139	0.103	0.064	0.050	0.015	0.005	0.002	0.000
Portsmouth	15.270	2.561	0.000	0.000	0.000	0.000	0.000	0.000	0.005	0.007	0.024	0.048	0.074	0.108	0.115	0.112	0.139	0.141	0.137	0.070	0.020	0.000	0.000
Blackburn Rovers	15.623	2.306	0.000	0.000	0.000	0.000	0.000	0.001	0.001	0.004	0.017	0.032	0.045	0.088	0.110	0.132	0.172	0.172	0.143	0.064	0.019	0.000	0.000
Sunderland	14.055	2.436	0.000	0.000	0.000	0.000	0.000	0.003	0.007	0.021	0.039	0.089	0.109	0.146	0.152	0.135	0.135	0.079	0.062	0.018	0.005	0.000	0.000
Hull City	14.049	2.550	0.000	0.000	0.000	0.000	0.000	0.000	0.001	0.013	0.030	0.051	0.077	0.110	0.129	0.122	0.163	0.126	0.086	0.066	0.025	0.001	0.000
Newcastle United	16.159	2.209	0.000	0.000	0.000	0.000	0.000	0.000	0.001	0.002	0.012	0.018	0.034	0.055	0.103	0.128	0.145	0.183	0.180	0.121	0.018	0.000	0.000
Middlesbrough	18.384	1.492	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.001	0.003	0.002	0.004	0.018	0.030	0.049	0.091	0.181	0.446	0.175	0.000	0.000
West Bromwich Albion	19.567	0.936	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.002	0.004	0.003	0.010	0.024	0.051	0.166	0.740	0.000	0.000

Received September 10, 2012; accepted October 21, 2013.

Francisco Louzada
Instituto de Ciências Matemáticas e de Computação
Universidade de São Paulo
Avenida Trabalhador São-carlense 400 - Centro
CEP: 13566-590 - São Carlos-SP, Brazil
Fone: 55 (16) 3373-6614 - Fax: 55 (16) 3371-2238
Email: louzada@icmc.usp.br

Adriano Kamimura Suzuki
Instituto de Ciências Matemáticas e de Computação
Universidade de São Paulo
Avenida Trabalhador São-carlense 400 - Centro
CEP: 13566-590 - São Carlos-SP, Brazil
Fone: 55 (16) 3373-9149 - Fax: 55 (16) 3371-2238
E-mail: suzuki@icmc.usp.br

Luis Ernesto Bueno Salasar
Departamento de Estatística
Universidade Federal de São Carlos
Rodovia Washington Luiz, km 235
CEP: 13565-905 - São Carlos-SP, Brazil
Fone/Fax: (16) 3351-8292
Email: luis.salasar@gmail.com