Fakultät für Wirtschaftswissenschaften
Facoltà di Economia
Faculty of Economics and Management

BEMPS –
Bozen Economics & Management
Paper Series

NO 72/ 2020

Dynamic Bayesian forecasting of English Premier League match results with the Skellam distribution

Robert C. Smit, Francesco Ravazzolo,
Luca Rossini

Dynamic Bayesian forecasting of English Premier League match results with the Skellam distribution*

Robert C. Smit† Francesco Ravazzolo‡ Luca Rossini§

May 10, 2021

Abstract

Due to teams trading players and/or changing their manager and players getting either in or out of injuries and other factors the abilities of teams to score or to prevent an opponent from scoring changes throughout the season in most if not all sports and association football is no exception. As such, we developed a dynamic model based on the Poisson difference distribution, called Skellam, where the scoring abilities are changing over time and are different across teams. The model is developed in a Bayesian framework and is fitted using the Stan modelling language. In this paper, we introduce a unique method used to handle promotion and relegation within the league. The model uses 3 different seasons and the forecasting ability has been measured and validated on the 2018-2019 English Premier League season. As a result, the model predicts the outcome of the matches correctly about 60% of the time. Moreover, we find that the model under-performs somewhat with the best performing, worst performing teams and some of the promotion teams, which could be attributed to both the fact that the season was an outlier in regards to performance for these teams and to the possibility that the hierarchical model may have caused shrinkage.

JEL codes: C11, L83, Z20.

Keyword: Bayesian hierarchical models; dynamic models; English Premier League; football data; Skellam distribution.

1 Introduction

Forecasting football matches is very popular. In sports betting it is common to simply indicate if one expects a team to win, lose or draw. Sports betting sites give the opportunity to bet on the outcomes of football matches on a weekly basis which gives an incentive to develop statistical models to help make informed bets. Many factors can determine the eventual result of a match, including the attacking and defensive strength of a team, the home ground advantage (when applicable) and specific match events. There are several ways to try to

*The authors gratefully acknowledge Andrew Manderson, Kevin Murray and Berwin Turlach for their useful feedback. This paper is part of the research activities at the Centre for Applied Macroeconomics and Commodity Prices (CAMP) at the BI Norwegian Business School.

†Vlammrs, The Netherlands r.thesmit@gmail.com

‡Free University of Bozen-Bolzano, Italy, CAMP, BI Norwegian Business School, Norway, and RCEA. francesco.ravazzolo@unibz.it

§University of Milan, Italy. l.rossini@qmul.ac.uk

model these factors to get an informed bet. Koopman and Lit (2019) considers 3 possible observational variables on which forecast the results of the next match in terms of win, loss or draw. The first variable is two-dimensional and consists of the numbers of goals scored by the two opposing teams. The second variable is the difference between the number of goals scored and the third is simply an indicator of a win, loss or draw. Clearly the informational content of these variables is decreasing. Besides the observational variable one can also categorize the different studies done in this field on the basis of the dynamic extension of the model used in the study. This extension can be either fully dynamic, "semi-dynamic", where one uses a method of weighted maximum likelihood estimations, or static if no dynamics were used. Combing these possible extensions, the literature can be organized into nine categories, and for each of these possibilities there are contributions extending over multiple fields such as statistics, econometrics and machine learning. As part of the introduction a discussion of the literature on forecasting football matches will follow.

Most of the studies on the modelling and forecasting of matches focuses on the first variable described above, where the two observations of the number of goals scored per teams are assumed to come from a bivariate distribution. Formally the probability of a given match outcome is given as follows $P(X = x, Y = y)$, for $x, y \in \mathbb{N}_0$, where X and Y are the numbers of goals scored by the home and away teams, respectively. The usual focus with forecasting is the probabilities of a home win, draw or away win; these are the so-called *toto probabilities* which are given by $P(X > Y)$, $P(X = Y)$, and $P(X < Y)$, respectively. In the late-1970s the binomial or negative binomial was suggested as the distributional form by Pollard et al. (1977). Since then however it has been widely accepted that the Poisson distribution is the more suitable model for association football results. Since using the Poisson distribution, the parameters of the distribution can be expressed as a function of how strong the competing teams' attacks and defences are. The original proposer of this procedure was Maher (1982), who expressed the team-specific strengths of attack and defence as a product of two independent Poisson distributions. In the paper, these parameters were constructed as the product of the attacking strength of one team and the weakness of the defense of the the other, which both were estimated statically.

Dixon and Coles (1997) considered the same distribution, and introduced a dependence parameter for the match results 0-0, 1-0. 0-1 and 1-1. Another proposal made by Dixon and Coles (1997) was a weighting function to put less weight on the likelihood contributions of observations from the more distant past meaning that the model was semi-dynamic. Crowder et al. (2002) formulated the model of Dixon and Coles (1997) as a non-Gaussian state space model with time-varying strengths for attack and defence, then developed approximating methods for both the parameter estimation and the signal extraction, since according to them the exact analysis was too expensive to compute. In practice, Crowder et al. (2002) extended the model to be dynamic instead of semi-dynamic. Karlis and Ntzoufras (2003) showed while also using the bivariate Poisson distribution that the introducing a parameter for the dependence between the goals scored by the two competing teams leads to a more accurate outcome prediction of a draw as a result than adding these dynamics. Rue and Salvesen (2000) incorporated the same framework as Dixon and Coles (1997) and develop a Dynamic Generalized Linear Model (DGLM) which is analysed continuously over time by Markov chain Monte Carlo methods. Owen (2011) also proposed using DGLMs and adopted Markov chain Monte Carlo (MCMC) simulation methods to derive the parameter estimates and predictive probabilities of match outcomes. Certain model specifications and identifiability constraints that were implemented in Owen (2011) were also used in this paper and will be discussed later.Recently, Koopman and Lit (2015) used a non-Gaussian state space framework based on the bivariate Poisson model with stochastically time-varying attack and defence strengths, and with some of the above extensions to effectively analyse weekly match results.

The next most popular of the three types of observations is modelling the toto probabilities directly. The modelling of the match results rather than the score or goal difference should lead to more parsimonious models and a simpler estimation procedure. With regards to static models, Koning (2000) investigated the balance in competition in Dutch professional football with an ordered probit model while having static team strengths. Other static models using the toto probability as their observational variable are Forrest and Simmons (1999) and Goddard and Asimakopoulos (2004) who proposed an ordered logit model and an ordered probit regression model respectively. An early contributor to this section of the forecasting football results literature was actually a dynamic model and is made by Fahrmeir and Tutz (1994) who introduced an ordered logit non-Gaussian state space model that incorporated random walks for the team strengths. The parameter estimation is then done by the Kalman filter and recursive posterior mode estimation methods. Knorr-Held (2000) applied their dynamic cumulative link model based on the extended Kalman filter and smoother to German Bundesliga data. Held and Vollnhals (2005) adopted the model from Knorr-Held (2000) for evaluating the comparative strengths of the football teams in the 5 top European leagues and linked them through the European competitions. Hvattum and Arntzen (2010) had a different proposal using ELO ratings updating over time as their way of representing team strengths with an ordered logit model. The ELO rating system was originally developed to assess the strength of chess players (Elo (1978)), but has been widely adopted in various other sports, including association football (Buchdahl (2003)).

The least popular category is the difference in the number of goals scored in a match which can also be seen as a team's margin of victory. In this category we let $Z = X - Y$ be the difference between the numbers of goals scored by the home team X and the away team Y with $Z = \mathbb{Z}$. By modelling Z, we can give the toto probability for a home win, draw and away win by $P(Z > 0)$, $P(Z = 0)$ and $P(Z < 0)$ respectively. The disadvantage that modelling the goal difference has over modelling the goals is that you will lose some information, since for example, the pairs $(X = 0, Y = 1)$ and $(X = 2, Y = 3)$ produce the same Z value. The advantage that comes along with that though is that you need a smaller number of summations in order to obtain the toto probabilities from Z compared to the pair (X, Y). Even with given the advantage using the one variable has over the other and vice versa it might not be immediately clear what the overall effect is of modelling Z over (X, Y). The reasoning behind it is that the accumulation of modelling error could be smaller, since a smaller number of probability components are being summed compared to modelling for the goals of both teams. A model for the differences between the number of goals scored in football matches is provided by Karlis and Ntzoufras (2009), who introduced the Skellam distribution for analysing match results. This distribution was originally derived by Skellam (1946) as the difference of 2 independent Poisson distributions. Karlis and Ntzoufras (2009) show however that the independence is not strictly necessary and also that the Poisson assumption for the variables is not needed. This was then later combined in Manderson et al. (2018) to forecast results in a different sport namely Australian Rules Football with its league the Australian Football League (AFL).

Lastly Koopman and Lit (2019) needs to be mentioned once again as it did not only make the consideration to categorize the literature into the 3 different observational variables it also contributed to all of them. Koopman and Lit (2019) developed a new dynamic multivariate model for the analysis and forecasting of football match results where the dynamic extensions are based on the class of score-driven models in which the time-varying coefficients are updated using autoregressive processes. This process is driven by the score of the conditional observation probability density function. This was then used to investigate which of the three model classes performs best in forecasting the toto probabilities in the next round of the competition. In the study dynamic models generally outperformed the static models and the dynamic bivariate Poisson model (goal scoring) and the dynamic Skellam model (goal difference) turned out to outperform the ordered probit model (toto probability). The conclusion was that the merging of data (from two counts, to the difference, to the sign of the difference) leads to decreasing forecasting performance and that their dynamic model outperformed the benchmark static models both in precision and betting losses.

Seeing as there is not as much research done in the goal difference category, this study will look into expanding this particular category. We propose a modelling framework similar to that of Manderson et al. (2018), however the changes that were made specifically to accommodate the nature of Australian Rules Football are redirected to better suit (association) football matches. Besides the scoring there is also another difference between the AFL and most association football leagues. While the AFL does not have a system of promotion relegation, most association football leagues do it such as the English Premier League. In order to accommodate for that most researchers choose to create models for each club and only for the years that they are a part of the researched league, however we have chosen not to go along this same path. Instead each season the estimated model of the relegated sides will be given to the newly promoted sides. How this works and why this was done will be further discussed in section 3.1.

From applying this model to the English Premier League 2018-2019 season we find that the model can generally accurately predict who wins or that the result will be draw just over 60% of the time. The model seems to struggle with correctly forecasting the results of the best and worst performing teams as well as some of the promoted teams this might be due to the performance of these teams being historical outliers.

The remainder of the paper is structured as follows: First the framework and details of the model will be discussed in section 2 which includes a short description of the Skellam distribution, followed by the construction of the dynamic Bayesian hierarchical model and the identifiability constraints. Thereafter in section 3 the data description will be given as well as how the data was handled such as the aforementioned handling of promotion relegation. In Section 4 the application of the methodology to data from the English Premier League 2018-2019 will be described and some results will be provided. Lastly in section 5 the conclusions are provided.

2 Model framework

The core of the model consists of the Skellam distribution. Skellam (1946) defines the distribution as the difference between two independent Poisson random variables. Consider two independent Poisson random variables X and Y, with parameters λ_x and λ_y respectively. Then the probability mass function of $Z = X - Y$ is as follows:

$$f(z|\lambda_x, \lambda_y) = e^{-(\lambda_x+\lambda_y)} \left(\frac{\lambda_x}{\lambda_y}\right)^{z/2} I_z\left(2\sqrt{\lambda_x \lambda_y}\right).$$

In this function, $I_z(x)$ is the modified Bessel function of the first kind and the score difference $z \in \mathbb{Z}$. In Figure 1 an example is shown of the Skellam distribution related to the full sample of the English Premier League seasons. The modified Bessel function of the first kind can be generally defined as

$$I_z(x) = \frac{1}{2\pi i} \oint e^{(x/2)(t+1/t} t^{-z-1} dt,$$

which can be simplified with z is an integer (see Abramowitz and Stegun (1965), Weisstein (2002)) to:

$$I_z(x) = \frac{1}{\pi} \int_0^{\pi} e^{x cos\theta} cos(z\theta) d\theta.$$

The mean and variance of Z are given by

$$\mathbb{E}(Z) = \lambda_1 - \lambda_2, \qquad \mathbb{V}ar(Z) = \lambda_1 + \lambda_2.$$

As a remark, if there are any possible correlations between the team scores then the variable $Z = X - Y$ will still follow a Skellam distribution. For example if we would have $X = X_1 + X_0$ and $Y = X_2 + X_0$ where all the X_is ($i = 0, 1, 2$) are independent Poisson distributed random variables (Koopman and Lit (2015)), then Z would be equal to $X - Y = X_1 - X_2$ which is still the difference between two independent Poisson distributed random variables and thus would still follow the Skellam distribution. For this reason, weather, time of play, pitch conditions, among several other factors influencing both teams equally, can be appropriately modelled through an X_0 as in the example before. Hence, these factors do not need to be explicitly

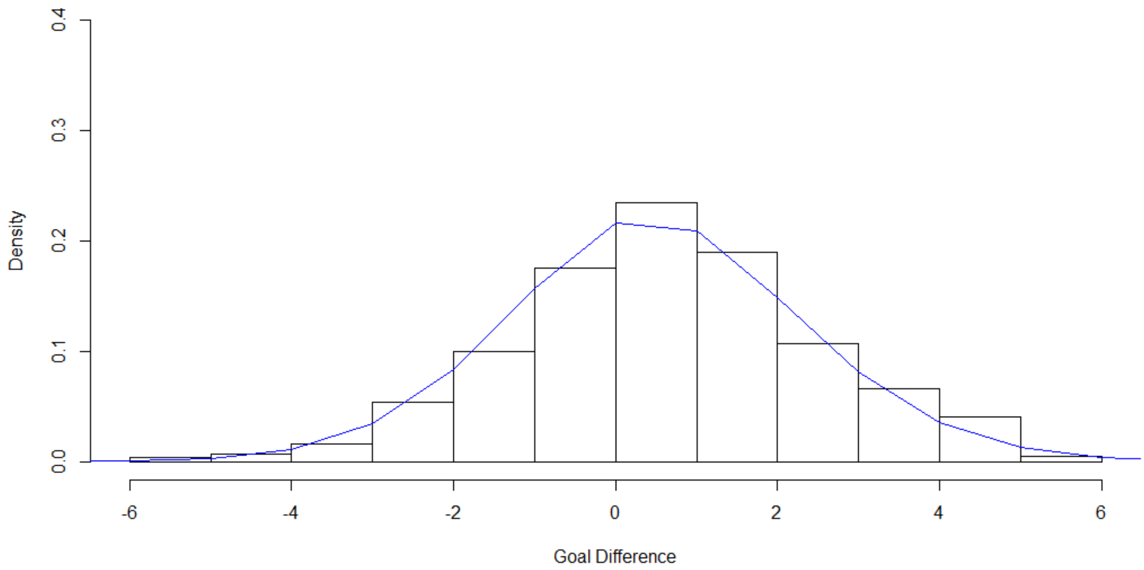

Figure 1: The goal difference for the 2016-2019 English premier League seasons, with Skellam distributions fitted for the Top 5 ranked teams at the end of the 2018-2019 season. The parameter values of the Skellam distributions are $\lambda_1 = 2$ and $\lambda_2 = 1.5$.

modeled for.

2.1 Dynamic structure

Since it can be reasonably assumed that the strength of teams depends on their opponent and changes over time, the parameters that determine the goal difference λ_x and λ_y similarly need to change over time and depending on their opponent. In other words, they need to be dynamic. Let teams i and j play each other at time t where team i plays in their home stadium. The combination of how potent the offence is of a team and how sound the defence is of the opposing team should determine the scoring ability of a team. Similarly to Manderson et al. (2018), there is also a home ground advantage included as another factor within the model as this is also a well-known factor within association football. Let $\alpha_{i,t}$ and $\beta_{i,t}$ be the attacking and defensive strength of home team i at time t and let δ be the home ground advantage for all teams over all time points. Using these definitions in relation to λ_1 and λ_2, the log link functions can be defined as:

$$\lambda_{1,i,j,t} = exp(\delta + \alpha_{i,t} - \beta_{j,t}), \qquad \lambda_{2,i,j,t} = exp(\alpha_{j,t} - \beta_{i,t}),$$

where $\lambda_{1,i,j,t}$ is the scoring ability of the home team and $\lambda_{2,i,j,t}$ is the scoring ability of the away team and $i \neq j$ and t is the time parameter. We assume that each team has its own attacking and defensive strength for each match and together this represents the performance of that team for that match. The reasoning for using the δ at all here is that if we look over all the considered seasons the home team on average scores 1.5658 goals with a standard deviation of 1.3192 which is slightly more than away teams who score 1.2009 goals with a standard deviation of 1.1952 as you can see in Table A.4 in section A.1. As in Manderson et al. (2018) we argue that a different δ parameter per team would be an insufficient specification. As the strength of a team changes, the home advantage might change accordingly. Hence the parameters would also need to change over time and depend on their opponent. This would lead to a large number of extra parameters. To avoid overfitting the model, you would then also need to put into place heavy regulations on these extra parameters. This would make the model unnecessarily complex. Another note should be made of the fact that in contrast to Manderson et al. (2018) there is no λ_3 and λ_4 for the behind difference. Seeing as these are specific to the sport of Australian Rules Football and do not have a similar counterpart in European Football Leagues.

With the use of autoregressive processes similar to the once used in Manderson et al. (2018) on both the attacking and defensive strengths we can create variation in them over time. Explicitly:

$$\alpha_{i,t} = \mu_{\alpha,i} + \phi_\alpha \alpha_{i,t-1} + \varepsilon_{\alpha,i,t},$$
$$\beta_{i,t} = \mu_{\beta,i} + \phi_\beta \beta_{i,t-1} + \varepsilon_{\beta,i,t},$$

where the autoregressive means are denoted by $\mu_{\alpha,i}$ and $\mu_{\beta,i}$, the coefficients are denoted as ϕ_α and ϕ_β and lastly $\epsilon_{\alpha,i,t}$ and $\epsilon_{\beta,i,t}$ are error terms with a normal distribution. Comparably to Koopman and Lit (2015) and Manderson et al. (2018), to reduce the number of parameters for precise inference, the assumption is made that the autoregressive coefficients are constant for each team. Since it is expected that the possible shift in performance from game to game to be similar between teams. The error terms are assumed to be independently and normally distributed:

$$\varepsilon_{k,i,t} \sim N(0, \sigma_k^2), \qquad k = \alpha, \beta,$$

and are also assumed to be the same for all teams attacking and defensive processes.

2.2 Bayesian specification

For the last part of the model description we need to specify the priors. Notably priors are by definition not objective and are considerably influenced by other implementations of similar models. On the standard deviation we place a Gamma[1] prior similarly to Baio and Blangiardo (2010), as

$$\sigma_k \sim \Gamma(0.1, 0.1), \qquad k = \alpha, \beta.$$

On the ϕ parameters of the autoregressive process, we place the following prior:

$$\phi_k \sim \Gamma(9, 10), \qquad k = \alpha, \beta,$$

[1]For the used parametrization $x \sim \Gamma(a, b)$ has $E[x] = a/b$ and $Var[x] = a/b^2$

as the estimate for ϕ is expected to be high. Another argument for this being that Koopman and Lit (2015) had high estimated values for their Premier League data. As mentioned by Manderson et al. (2018) an added benefit to using a prior with large estimates is that any potential numerical problems that we might have come across from ϕ estimates that are nearing zero are avoided. As Manderson et al. (2018) also mentions as with their specification, this specification might be perceived as being quite informative, but we also have a data set which is quite large with several distinct time series. Meaning that the posterior estimate for ϕ can easily still take on values that belong to intervals with low probability.

2.3 Constraints for precise inference and its consequences

Similarly to Manderson et al. (2018), we encounter problems because of the structure of the time series. There are a large number of latent variables. This is problematic since it can create difficulties and inconsistencies with the fitting of the model. In order to make sure that every time the model is fitted it ends up at the same appropriate model each time or in other words to ensure precise inference (or identifiability) some constraints need to be added. For each round in the data set, the parameters $(\alpha_{i,t}, \beta_{i,t})$ have to be considered for each team. Together with the offensive and defensive autoregressive mean of each team (μ_α, μ_β), the autoregressive coefficients $(\phi_\alpha, \phi_\beta)$, the home ground advantage (δ) and other latent variables this results into a lot of parameters and latent variables with the used data set.

As with Manderson et al. (2018), how good the two teams are comparatively is the only important value in the end. Thus nothing changes for the effectiveness of the model as long as the relative difference between the autoregressive estimates remains equal, thus how much stronger a team's attack strength is compared to its opponent is not being altered for example. As such we can also use a sum-to-zero constraint to all the autoregressive parameters without affecting the model negatively:

$$\sum_{i=1}^{n} \alpha_{i,t} = 0, \qquad \sum_{i=1}^{n} \beta_{i,t} = 0, \tag{1}$$

for all $t \in \{1, ..., T\}$. T is the total amount of rounds in the data set and n is the number of teams in the Premier league which for all considered seasons is 20. In section 3 the specifics of these values are discussed. Again using the same method as in Manderson et al. (2018) the constraints are implemented by defining the nth parameter to be:

$$\alpha_{n,t} = -\sum_{i=1}^{n-1} \alpha_{i,t}, \qquad \beta_{n,t} = -\sum_{i=1}^{n-1} \beta_{i,t},$$

for a given time t.

However as Manderson et al. (2018) also pointed out this approach leads to the estimates for the nth parameters to be considerably higher or lower than any of the other parameter estimates at that given time. This might be because the nth parameter estimate being equal to the negative sum of the other estimates is only fine if the prior on the parameters (here on the first n-1) does not depend on the hyperpriors, as the hyperpriors should be informed by all n parameters, and not just n-1 as discussed in Kruschke (2011). However the prior on the parameters with this time series does not depend on the hyperpriors. A more in-depth discussion can be found in Knorr-Held (2000).

By using the same idea as Manderson et al. (2018) which itself came from Knorr-Held (2000) and Owen (2011) this numerical irregularity and computational inefficiency can be corrected. If the parameters $\boldsymbol{\alpha}_i$ for $i \in \{1, \dots, n\}$ are considered then the following multivariate normal distribution can be formed showing its relation to the past:

$$\boldsymbol{\alpha}_t \sim \mathrm{N}(\phi_\alpha \boldsymbol{\alpha}_{t-1} + \boldsymbol{\mu}_\alpha, \mathbf{D})$$

where $\boldsymbol{\alpha}_t = (\alpha_{1t}, \dots, \alpha_{nt})$ is a vector containing the attacking strength of teams 1 through n at time t, $\boldsymbol{\mu}_\alpha$ is a vector of autoregressive means for the α process and $\mathbf{D}$ is a diagonal matrix of σ_α^2's. Naturally the initial values of the time series would be $\alpha_1 \sim \mathrm{N}(\phi_a \boldsymbol{\mu}_\alpha, \mathbf{D})$ where $\boldsymbol{\mu}_\alpha$ is still a vector of autoregressive means for the α process. There is an implicit condition for the sum of elements of $\boldsymbol{\mu}_\alpha$ to be zero as the sum-to-zero constraints need to hold for all time points and $t = 1$ is not excluded from that. Additionally, in Owen (2011) it is shown that for equation (1) to hold the covariance matrix $\mathbf{D}$ must altered to be of the following form:

$$\mathbf{R} = \frac{n\sigma^2}{(n-1)} \Big(\mathbf{I}_n - \frac{1}{n} \mathbf{1}_n \mathbf{1}_n^{\mathrm{T}} \Big),$$

where n is the amount of teams in the league, $\mathbf{I}_n$ is an identity matrix of $(n \times n)$ and $\mathbf{1}_n$ is a vector of ones $(n \times 1)$. As mentioned by Manderson et al. (2018) with this form of $\mathbf{R}$ one faces other problems, since $\mathbf{R}$ is not of full rank and thus has no inverse which makes it impossible for STAN to sample from the multivariate normal distribution directly where $\mathbf{R}$ is used for the covariance structure. Instead we will opt for the same sampling method as shown by Manderson et al. (2018). The sampling will be done from the following $(n-1)$ dimensional multivariate normal distribution as it has an appropriate mean and covariance structure:

$$\mathbf{W} = \frac{n\sigma^2}{(n-1)} \Big(\mathbf{I}_{n-1} + \mathbf{1}_{n-1} \mathbf{1}_{n-1}^{\mathrm{T}} \Big),$$

hereafter these variates consistent with Manderson et al. (2018) will be referred to as the *raw* parameters. Transforming these *raw* parameters into the parameters of interest is done by the following $n \times (n-1)$ matrix:

$$J = \begin{pmatrix} \mathbf{I}_{n-1} - \frac{1}{n} \mathbf{1}_{n-1} \mathbf{1}_{n-1}^{\mathrm{T}} \\ \\ -\frac{1}{n} \mathbf{1}_{n-1}^{\mathrm{T}} \end{pmatrix},$$

the structure of J implements the desired centering constraints. To show what this entails practically take the $(n-1)$ dimensional *raw* means of all the autoregressive processes:

$$\mu_{\alpha,raw} \sim \mathrm{N}(\mathbf{0}, \sigma_\mu^2 \mathbf{W}), \qquad \mu_{\beta,raw} \sim \mathrm{N}(\mathbf{0}, \sigma_\mu^2 \mathbf{W}).$$

The *raw* means are then transformed by $\mathbf{J}$ into the parameters used for inference:

$$\mu_\alpha = \mathbf{J}\mu_{\alpha,raw}, \qquad \mu_\beta = \mathbf{J}\mu_{\beta,raw}.$$

Through this same process one can find the initial values for the team strength parameters, which are the following zero mean distributions:

$$\alpha_{t,raw} \sim \mathrm{N}(0, \sigma_\alpha^2 \mathbf{W}), \qquad \beta_{t,raw} \sim \mathrm{N}(0, \sigma_\beta^2 \mathbf{W}).$$

With those the initial values can be found:

$$\alpha_1 = \mu_\alpha + \frac{\sigma_\alpha^2}{\sqrt{1-\phi_\alpha^2}}\alpha_{1,raw}, \qquad \beta_1 = \mu_\beta + \frac{\sigma_\beta^2}{\sqrt{1-\phi_\beta^2}}\beta_{1,raw}.$$

Finally, starting from these initial values the following formulas for the 'attack' and 'defence' strengths can be formed by continuously going towards the next time point:

$$\begin{aligned}\alpha_{t+1} &= \phi_\alpha \alpha_t + \mu_\alpha + \sigma_\alpha^2 \mathbf{J}\alpha_{t+1,raw} \\ \beta_{t+1} &= \phi_\beta \beta_t + \mu_\beta + \sigma_\beta^2 \mathbf{J}\beta_{t+1,raw}\end{aligned}$$

To finish the hierarchical specification a prior distribution has to be chosen for σ_μ, which as Manderson et al. (2018) is chosen to be uniform:

$$\sigma_\mu \sim \mathbf{U}(0.005, 1).$$

The reason that this variance parameter bound is quite small is because there is limited variability in the value of μ. This reflects the fact that a single unexpected results only slightly changes the mean of the underlying autoregressive process. This is all similar to Manderson et al. (2018), although other Bayesian analyses do not allow this parameter to vary at all, and simply fix it at a particular values (see for example, Baio and Blangiardo (2010)).

In Figure 2 we present a graphical representation of this model, which is arranged as follows: the node in the bottom layer represents the data generating Skellam distribution, with the nodes in the second layer representing the parameters that govern this distribution, except for the black node in this layer which represents the home ground advantage. The nodes in the third and fourth layers relate to the time series structure, and lastly the nodes in the remaining layers represent the prior distributions required to complete the model specification.

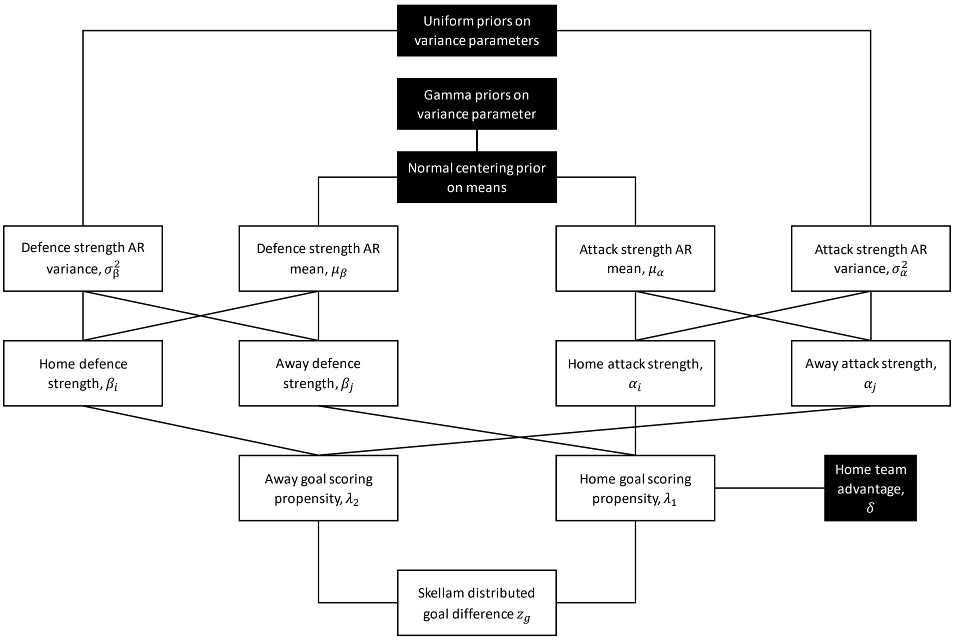

Figure 2: A graphical depiction of the structure of the model.

2.4 Inter-season variance

As most player purchases and training of players happens during the off season, a team can improve or decline from the end of a season to the start of the next season through several potential means like player purchases or sales or changes in management and training. As a time variable usually only indexes an occurrence it does not take into consideration this potentially substantial change between seasons.

To model for this change in performance during an off season can be considered difficult as a team's strategy in the off season may change considerably each year as well. Especially now that the Financial Fair Play (UEFA Financial Fair Play) was introduced in June 2010 which limits clubs over an assessment period of 3 years to not spend more money than they earn contrary to what would happen in the past were clubs would put themselves in high amounts of debt in order to get higher up or stay in the highest division. Thus there are a few data points associated with inter-seasonal change of performance, which are highly variable.

In the model, inter-season variance is handled by increasing the variance of the time series estimates at the start of each season. The variance of the rounds directly after are also increased but to a lesser extent. The assumption behind this is that the first round will show most of the information about a team's inter-seasonal change. For this we use the same exponentially decaying factor as Manderson et al. (2018) of $(\nu - 1)exp\{-(t - t_{ss})\} + 1$ such that:

$$\alpha_{i,t} \sim \left(\mu_i + \phi_a \alpha_{i,t-1}, \left[\left((\nu - 1)exp\{-(t - t_{ss})\} + 1\right)\sigma_a\right]^2\right),$$

where t_{ss} is the season start time point and ν is the inflation factor for the variance that is chosen.

3 Miscellaneous details & Data description

The English Premier League (EPL) has 20 teams (i.e. $n = 20$), which each season play each other 2 times once at their home ground and once at the other team's home ground for a total of 380 matches over 38 rounds each containing 10 matches. The data used here was from the seasons 2016-2017, 2017-2018 and the season 2018-2019 and can be found on http://www.football-data.co.uk/. Note that the seasons 2016-2017 and 2017-2018 were used as historical data to forecast the 2018-2019 season.

3.1 Promotion & Relegation

As with every football league a team receives 3 points for a win, one for a draw and zero for a loss. Also as in most football leagues at the end of the season a few teams go down to a lower league and are replaced by some teams from that lower league. In the case of the English Premier league 3 teams with the worst amount of points are relegated to the English Football League (EFL) Championship and replaced by the top two of the EFL Championship and the winner of the play-offs between the teams that finished in third through sixth in that league. Instead of for example opting to disregard the teams that are not in the data for all the considered years which would be 15 out of the 20 teams (25% of all teams in the league) we went with a different approach. Seeing that most of the time the newly promoted teams are deemed likely to immediately relegate back down. This inspired the choice to utilise the data and model of the relegated teams. To achieve this continuous use of the parameters every team is not only represented by its name, but is also given a team number. For the first season the assignment of the team numbers is simply done alphabetically. Thereafter however the non-relegated teams retain the same team number and the newly promoted sides will be given the number of a relegated team along the past season's final standings. We simplify this concept in Table 1. Here one can see the given team numbers per season with several teams highlighted. The teams in *italic* are the teams that in that particular season relegated to the EFL Championship. The teams written in bold are the teams that were newly promoted in that particular season.As you can clearly see these teams are not added alphabetically.

The replacement is done as follows: At the beginning of the EPL season last season's champion, runner-up and play-off winner of the EFL Championship will be assigned the parameters that the numbers 18, 19 and 20 had at the end of the previous season, respectively. Between the 2016-2017 and 2017-2018 seasons means the following: Hull, Middlesbrough and Sunderland were 18th, 19th and 20th, respectively in the final standings of the Premier League for the 2016-2017 season while Newcastle, Brighton and Huddersfield were 1st, 2nd and the playoff winner, respectively in the EFL Championship. As logic would dictate the best of the relegated teams would be Hull and it's estimated model is past down to Newcastle who would be the best of the promoted teams. The same goes for the second best Middlesbrough and Brighton and the worst of the three Sunderland and Huddersfield. This process is then repeated between seasons 2017-2018 and 2018-2019.

Although as you might notice only 2 out of the 9 teams those being Cardiff and Fulham in the 2018-2019 season that were newly promoted to the Premier league went immediately back down to the EFL Championship in the considered seasons, however this does not necessarily have to negatively impact the predictive ability of the model, since the parameters change over time according to the actual results. Thus if a newly promoted team performs very well the model will adjust to this. It should still aid the performance as supposed to alternative of using the quite uninformative prior since this can both severely underestimate or overestimate the performance of these newly promoted teams.

Team Number	2016-2017	2017-2018	2018-2019
1	Arsenal	Arsenal	Arsenal
2	Bournemouth	Bournemouth	Bournemouth
3	Burnley	Burnley	Burnley
4	Chelsea	Chelsea	Chelsea
5	Crystal Palace	Crystal Palace	Crystal Palace
6	Everton	Everton	Everton
7	*Hull*	**Newcastle**	Newcastle
8	Leicester	Leicester	Leicester
9	Liverpool	Liverpool	Liverpool
10	Man City	Man City	Man City
11	Man United	Man United	Man United
12	*Middlesbrough*	**Brighton**	Brighton
13	Southampton	Southampton	Southampton
14	Stoke	*Stoke*	***Cardiff***
15	*Sunderland*	**Huddersfield**	*Huddersfield*
16	Swansea	*Swansea*	**Wolves**
17	Tottenham	Tottenham	Tottenham
18	Watford	Watford	Watford
19	West Brom	*West Brom*	***Fulham***
20	West Ham	West Ham	West Ham

Table 1: Given team numbers per season of English Premier League. The bold teams are the newly promoted teams of that particular season and the *italic* teams are the teams that relegate that season.

3.2 Data adjustments

Besides the handling of the promotion and relegation there is also an issue regarding the rounds in the English Premier League. As mentioned at the beginning of this section the 380 matches of the league are divided into 38 rounds of each 10 matches. However because of other competitions whose matches are not all known beforehand it is also not uncommon that certain matches need to be moved to another date. This means that they were not played during the rest of the round but at an earlier or later date. Seeing as the data set has all matches sorted on date without including the round of which they were a part of this has consequences for the workings of the model. Namely the consequence of these rescheduled matches is that the way the model is set up to predict one round at a time does not work properly, because there were rounds where certain teams would play more than one game and other teams were not in that round. This would mean that some teams can earn up to 6 points in one round and others 0. Which it was not set up to do naturally since a team should only play one game each round. Allowing for this would make it impossible to accurately compare the teams' performance over the whole season.

For clarity we will give an example: The match between Everton and Manchester City in the 2018/19 season was originally part of round 27 (22-24 Feb 2019), but was actually played in between rounds 25 and 26 (6 Feb 2019). This results that if we take every 10 matches as being a round the match between Everton and Manchester City would be part of round 26 instead of 27 and the last match in the data set of round 26 would be part of round 27 and this would trickle down the whole data set. With multiple rescheduled games this problem of course can become a much bigger problem.

There are a few ways we saw into resolving this issue. One was simply finding a data set which included the round the match was originally a part of. Sadly such a usable data set could not be found. Another solution is allowing different amounts of matches to be considered a round. Problem with this solution is that you are still allowing a team to either not to be a part of a round or play in it twice. Which would mean that you still have the same problem as before of having some teams earn a maximum of 6 and others no points during a round.

A third possible solution was to rework the entire model to instead of working on just a round by round basis to work on a round by round team basis. Meaning that you would estimate each team round by round. This would however take a lot of time to run since instead of having to simulate for 38 rounds one would have to simulate those 38 rounds 20 times for each team once, which comes down to 760 simulations.

The last and eventually chosen solution is to manually adjust the data sets for all season so that the games that were being played at another time will be with the rest of the originally scheduled round. This has as a benefit over the other solutions that there are still always ten matches each round and each team is just represented once, which means that you can still see and compare how teams perform over the whole season. There is still a downside to this solution with regards to the form of a team, however, as it is quite common that during certain periods the team performs above their normal level and other periods when the team performs under their usual level. Meaning that if you shift games from when they were actually played to a different time this might negatively affect these patterns that you would otherwise see or possibly your predictions on those matches might be slightly off. However it should not have a bad overall impact on the evaluation of a team's strength.

4 Results

4.1 Model fitting, estimations and predictions

The described model of Section 2 was coded in the STAN modelling language and fitted using the RSTAN interface. Specifically, the concern is the estimation of the binary results being either a win or a loss, as well as the win, draw and loss (WDL) probabilities for the 2018-2019 English Premier League season. The data from 2016-2017 season to 2018-2019 were used to both fit the model and predict the results. Each round took a few hours to fit from home with acceptable convergence, which means that the model could be used round to round forecasting setting.

As Manderson et al. (2018) we took the most important parameters for a stable model those being $\alpha_{i,t}$ and $\beta_{i,t}$ in our case ($\forall\ i \in \{1, \ldots, n\}$ and $\forall\ t \in \{1, \ldots, T\}$) and looked at traceplots and density estimates to assess convergence of the MCMC chains. In Figure 3 we show the traceplots and density estimates of the parameters $\alpha_{1,38}$ and $\beta_{1,38}$ as an typical example and they do seem to converge. Other examples for the top 5 teams of the 2018-2019 season at three different time points, specifically the first round, the mid-season point (round 19) and at the end-of-season (round 38), can be found in appendix A.3. Furthermore in Figure 4 and

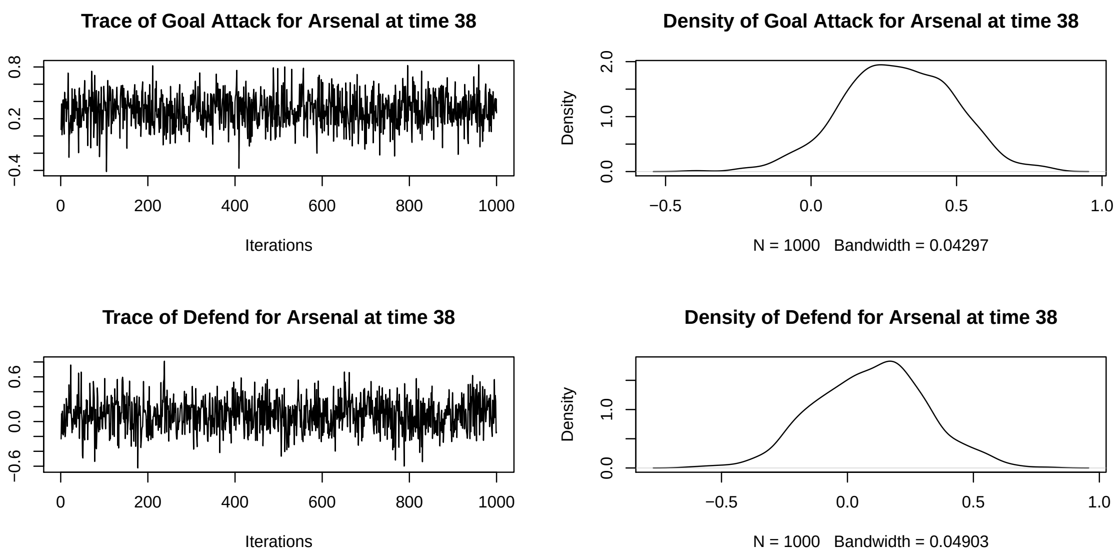

Figure 3: A traceplot and density estimate of the Goal Attack (top) and Defence parameter (bottom) for Arsenal at time point 38 generated with 1.000 iterations.

5 you can see an example of the variation of the attacking strength and defence strength over the different rounds which formally can be seen as the mean of the posterior distribution also known as the posterior mean. In order to best compare the teams in these figures only the top 5 ranked teams in the 2018-2019 season are shown. These teams ranked in the following order in the final standings of the 2018-2019 season: Manchester City, Liverpool, Chelsea, Tottenham and Arsenal. The posterior means for the strength of all teams over the different rounds can

be found in appendix A.2. In these figures we see that for the most part of the sample for both the attacking and defence strength do not change in a major way after initialisation. The initialisation is generally quite different from the values that the posterior means hover around, because of the uninformative priors. However because of the dynamic elements the posterior means will come closer to what the "actual" value might be. But besides the initialisation there are still some noteworthy variations. For defence strength the most notable larger variations happen for Arsenal and Chelsea at the beginning of the second season, the 2017-2018 season. This can be explained through a combination of form of the team and the inter-season variance that decays over several rounds. As a general rule most variations could be attributed to either or both of these reasons. For the attacking strength only Manchester City seem to drastically vary mostly in the 2017-2018 season, but if we take into consideration the graph of Manchester City goals scored and conceded in Figure A.1 we can argue that goals scored during the 2017-2018 season greatly vary as well particularly in the beginning of the 2017-2018 season. This combined with the inter-season variance can explain why there is a huge spike in the attacking strength of Manchester City at the beginning of the 2017-2018 season as well as during all seasons.

Besides the individual variations in strength both in defensive and attacking sense another note can be made about the differences between teams. For attacking strength Manchester City over these particular seasons are constantly above the rest. There are times that it is close, but Manchester City never gets passed in attacking strength in the model. The other four teams displayed in Figure 4 are far closer together and it constantly changes which team has an higher attacking strength. For defence strength it is a slightly different story however, Manchester City also has the best defence strength over all the researched seasons only falling below other teams about a handful of times, however, the difference between itself and two other teams of the "Top 5" are far smaller than with the attacking strength. The two teams closer to Manchester City with regards to Defence strength are Liverpool and Tottenham, who are quite close to each other and in the first two seasons seem to jump each other quite frequently only in the last season does Tottenham not pass Liverpool in regards to defence strength. Below these two teams are the last two teams of this "Top 5" Arsenal and Chelsea whose Defence strength sits a bit below that of Liverpool and Tottenham but Arsenal and Chelsea also constantly pass each other when it comes to the attacking strength in this model. Given that these figures show the strength of both the defence and offensive of the teams you could believe that you could already with some confidence give the relative ranking of these teams for all three seasons, however this does not work. If we look at Table 2 we can see that the relative rankings of these teams do not fall in completely fall in line with the combinations of offensive and defensive strength posterior means shown in Figure 4 and 5. This is because these graphs although based on previous results do not give you the actual results and even though if a team has an occasional unexpected loss or draw this does not severely impact their estimates for defence and attacking strength it might have a major impact on their chances of actually winning the league.

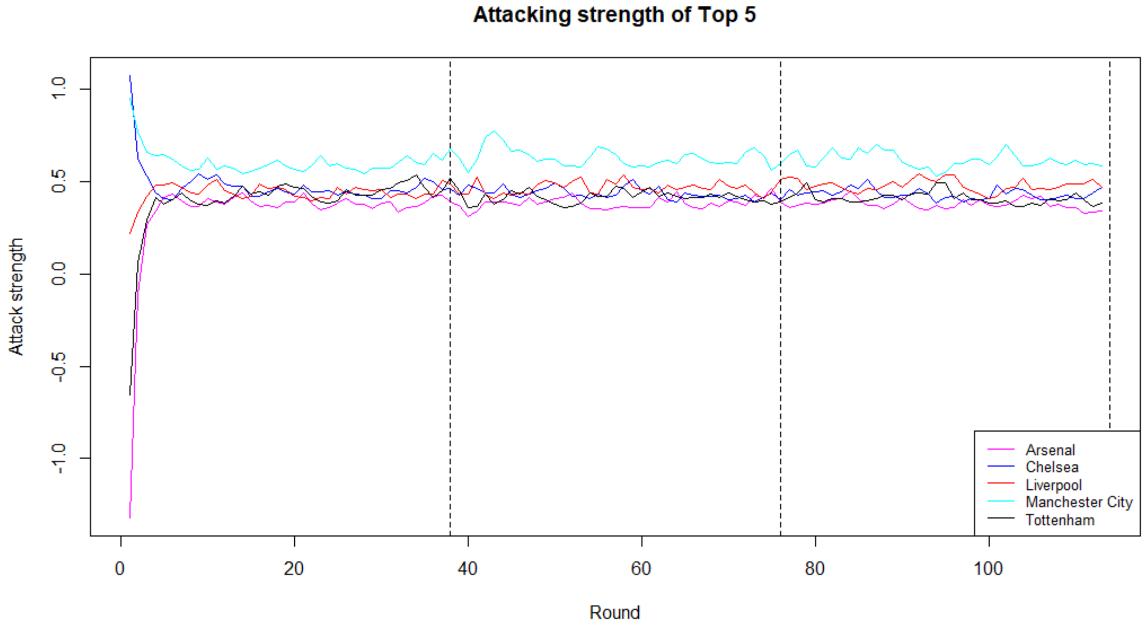

Figure 4: Posterior mean of the attacking strength from the 5 highest ranked teams (magenta for Arsenal, blue for Chelsea, red for Liverpool, light blue for Manchester City and black for Tottenham) in the 2018-2019 English Premier League season, where the vertical dotted lines indicate the end of a season.

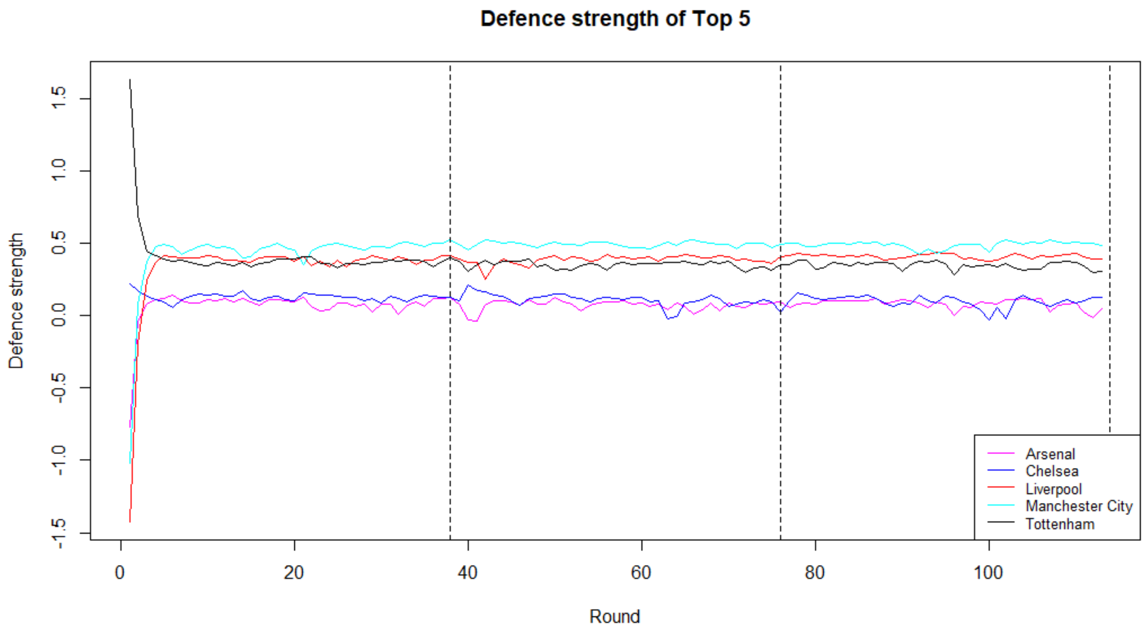

Figure 5: Posterior mean of the defence strength from the 5 highest ranked teams (magenta for Arsenal, blue for Chelsea, red for Liverpool, light blue for Manchester City and black for Tottenham) in the 2018-2019 English Premier League season, where the vertical dotted lines indicate the end of a season.

2016-2017	2017-2018	2018-2019
Chelsea	Manchester City	Manchester City
Tottenham	Tottenham	Liverpool
Manchester City	Liverpool	Chelsea
Liverpool	Chelsea	Tottenham
Arsenal	Arsenal	Arsenal

Table 2: Relative rankings in the final standings of the 2016-2017, 2017-2018 and 2018-2019 English Premier League season from the top 5 ranked teams of 2018-2019 season

4.2 Predictive results of the 2018-2019 English Premier League season and model validation

Predictions for match results were obtained by continuously going one-step ahead, meaning that all results from the previous rounds were used to predict the current round. The associated probabilities and estimated differences were generated by sampling from the posterior predictive distribution of the parameters. For practical reasons these predictions included also a predicted point difference, as well as the associated WDL probabilities. There were also prediction intervals constructed for the score differences of each match and a check whether each actual difference was contained within their corresponding interval.

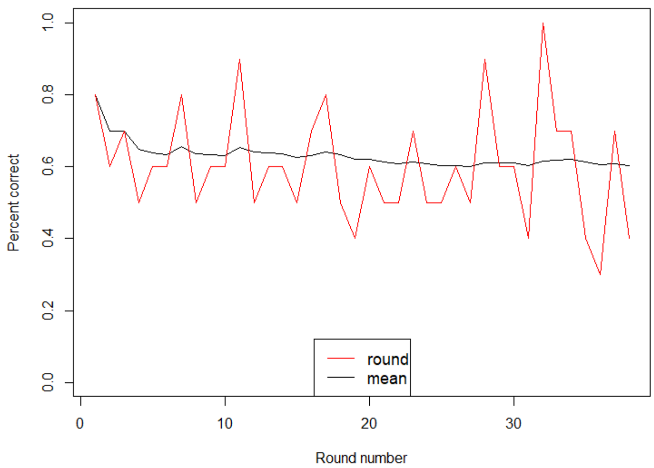

Figure 6: Predictive performance model over the 2018-2019 season round by round (red line). The black line is the fraction of the correctly predicted games over the season up to and including that point (e.g. point at round two is mean proportion of games correctly predicted in rounds 1 and 2).

Also as in Manderson et al. (2018) Figure 6 shows how accurate our predictions for the results were each round across the 2018-2019 season, where the red line shows the fraction of correctly predicted matches that round and the black line is the fraction of correct predictions up until then including the round itself or in other words the mean performance. The total number of correctly predicted matches was 229 out of a possible 380 which means that just over 60% of all games were predicted correctly. Over time there are generally periods the performance is above the mean and periods below it. What is immediately noticeable however is that after round 30 both the best and worst performance is reached. The best being round 32 where every match was predicted correctly, however 6 rounds later in round 36 the least amount of matches had the correct prediction with only 3 out of 10. A possible explanation for this big difference in performance might have to do that while some clubs are still playing for something whether that is winning the league, reaching either Champions League or Europa League qualification standings or avoiding relegation other teams might be complacent because they have already reached their goal or their goal is out of reach. However it is also possible that it may come from fatigue or other unknown factors which presented itself as random variation.

For the most part however the model predicts at least 50% of matches correctly. The only exceptions being the before mentioned end of the season, but also one round in the middle of the season more specifically round 19, which as it turns out was the round in the season 2018-2019 played on the 26th of December which is known in the United Kingdom as Boxing Day. In the UK there is a longstanding tradition to play on Boxing Day originally done to bumper audiences, since all workers had the day off. Now it is just a long rooted family orientated tradition.[2] However as you can imagine there is a lot more pressure from fans and for British players also family to perform well on Boxing Day making it likely that any team can either perform better or worse than they normally would. Following round 19 on Boxing Day there are 2 more rounds between Boxing Day and January 4th, where normally there is about one week between each round. Which could possibly lead to bad performances as well and although the model performs under the mean it does predict the results correctly about half of the time with around 16 games predicted correctly out of a possible 30 in rounds 20, 21 and 22 altogether.

As mentioned by Manderson et al. (2018) looking at the outcomes is definitely interesting for when one would want to use this model practically, it is not necessarily the best way to judge the performance of the model. For example, if the model predicts a goal difference of +1, but the model also has an estimated standard deviation of 3, and the actual result is either 0 or -1 than one would consider the prediction to be good and reasonable. However if you only look at the exact prediction of a win, loss or a draw then you would only see this prediction as incorrect. To have better insights in the model's performance, similarly to Manderson et al. (2018) a approximate prediction interval is constructed by using the estimations and standard deviations of the model. For the same reasons as with they had of having a large amount of samples and seemingly symmetric traceplots a normal interval can also be constructed. In Figure 7 you can see both the coverage probability estimated by the model and the actual coverage probability of the $100(1-\alpha)/2\%$ prediction intervals as a function of α. For all values of α it shows that the model is reasonably well behaved, since the difference between the actual coverage and nominal coverage not too large. One can also see that the actual coverage is generally slightly higher than the ideal one. Another validation method for the

[2] https://www.bbc.co.uk/bitesize/articles/z4y847h

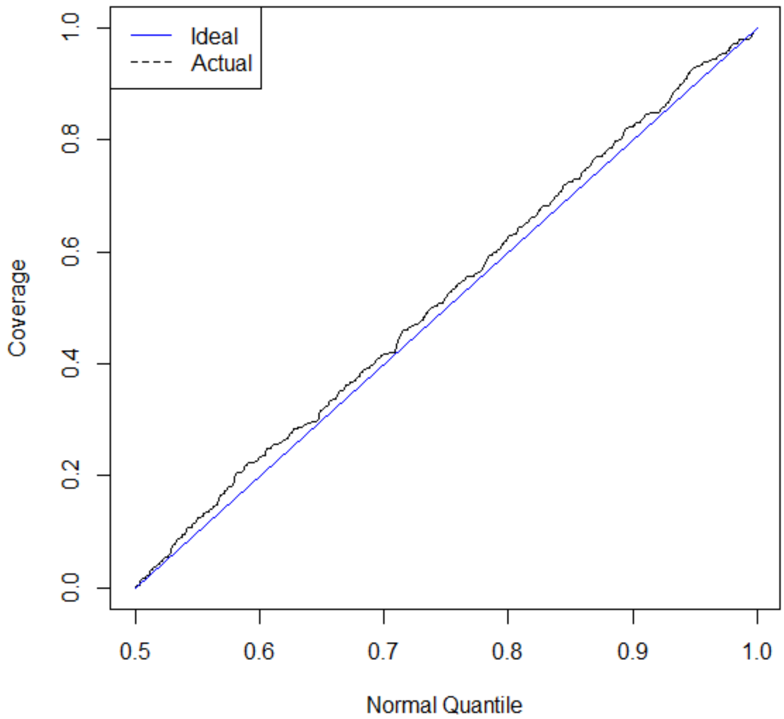

Figure 7: The coverage probability ($(1-\alpha)/2$) as a function of the normal quantile α used for the prediction interval construction. The blue and black lines show the nominal coverage and the actual coverage, respectively.

predictive capabilities of the model in particular the estimated WDL probabilities that were used was estimating the amount of points each team should have at each time point in the season, similarly to Manderson et al. (2018). By taking the estimated probabilities for each type of result and multiplying it by the amount of points that can be earned for such a result (which is as mentioned before three for a win, one for a draw and zero for a loss) gives the expected number of points in each match. With that we can map both the predicted points and the actual points for each team for the season round by round. The result of this can be seen in Figure 8. Note that for most teams the predicted and actual point totals are relatively close. There are a few exceptions to this. Liverpool is the highest ranked of these exceptions who scored 97 points compared to the predicted point total of around 78. This was however an historical performance, since this was the third-highest total in the history of the English top division and the most points scored by a team without winning the title.[3] Implying that it was unlikely to happen.

Wolves is another exception. The newly promoted team ended with more points than predicted as well in their case with 57 actual points where there were 43 predicted. This is however also

[3] https://www.theguardian.com/football/2019/may/12/liverpool-wolves-premier-league-match-report

quite an historic accomplishment seeing as they were ranked higher than any promoted side since Ipswich Town in 2000-2001.[4]

The last two teams that had a great discrepancy between their actual acquired points and the amount of points that the model predicted are two relegated teams namely Huddersfield and Fulham. Huddersfield finished the Premier League season at the bottom of the table with 16 points, one of the lowest totals ever recorded in the competition's history. According to some sources, their relegation can be blamed on their shoestring budget and mismanagement this season.[5] Fulham had no shoestring budget spending over £100m, but according to experts there was mismanagement which can be somewhat backed up by the statistics. Their first manager was fired after not getting win in seven matches, the second manager did not help improve Fulham however seeing as Fulham did not win a single league match over the winter period. The third manager did win three matches in the last month, which ended up making the difference between the predicted point total and actual points a bit smaller, however still significant.[6]

Another explanation for the worse performance shown by the validation method seen in Figure 8 is that the model, as is expected of a hierarchical model, causes shrinkage. Similarly to Manderson et al. (2018) the under-performing teams such as Huddersfield, Fulham and perhaps also Watford having their win rates over-predicted while higher performing teams like Manchester City, Liverpool and Wolves have their win rates under-predicted by the model. Unlike Manderson et al. (2018) however, there does not seem to be an apparent higher difficulty in predicting outcomes at the start of the season. Perhaps because unlike in their researched season of the AFL where teams unexpectedly won or lost games at the start of the season in the beginning of the 2018-2019 Premier League season the results were more as expected.

[4] https://www.itfc.co.uk/club/history/

[5] https://www.90min.com/posts/6368775-huddersfield-town-2018-19-review-end-of-season-report-card-for-the-terriers

[6] https://www.90min.com/posts/6368723-fulham-2018-19-review-end-of-season-report-card-for-the-cottagers

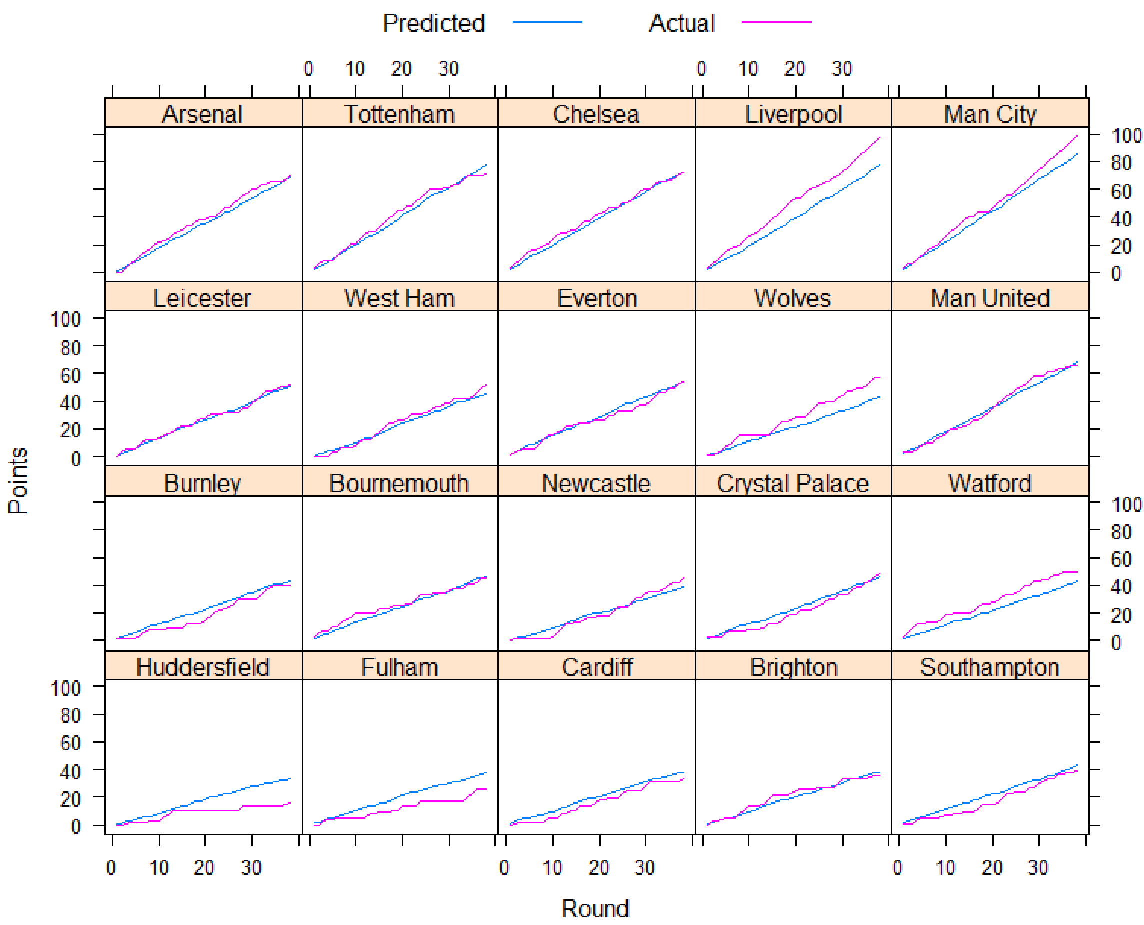

Figure 8: The predicted competition points (blue lines) and the actual total point (magenta lines) of each team over the 2018-2019 season round by round. The plots are sorted from right to left, top to bottom based on the actual standings at the end of the 2018-2019 season.

5 Conclusions

Currently there is less research done in using the difference in the amount of goals scored as the observational variable in attempting to forecast results of association football matches. This paper provides an approach to modelling football results using dynamic hierarchical Bayesian models similar to the model used by Manderson et al. (2018) for modelling AFL results which in turn was based on Baio and Blangiardo (2010) and Owen (2011). The models are based on the Skellam distribution and appropriate methods were applied to rectify problems in constraining parameters, model identifiability, scheduling of the match rounds and promotion & relegation. The final model is in the form of a Stan model, which enables the generation of predictive results and applied this model to data from the 2018-2019 English Premier League season, and achieved a comparable predictive performance to other soccer outcome models.

Future research can be done for the interesting 2019-2020 season. Besides it being a interesting season in general for several football-related reasons one could look at what the effects were of the midseason stoppage of matches being played, because of the Covid-19 pandemic. Potentially this could be another break in the data similar to the break between seasons and therefore potentially the inter-season variance method can also be applied to this break.

Furthermore as the newly promoted teams, best and worst performing teams had the worst predictive performance possibly these teams need to be identified in order to introduce some extra adjustments as a means to increase the predictive capabilities and also to reduce shrinkage issues. This might also aid into more conclusively being able to state whether or not the method used for handling the promotion & relegation where the promoted team adopts the predicted parameters of a relegated team is viable.

There can also be further extended upon this paper by validating the model through its viability in sports betting. What we have not read as being researched is not only testing which model performs best in regards of sports betting in predicting whether the result will be a home win, away win or a draw, but also testing whether the models predict either the exact score or the goal difference well enough so that it can increase the success in sports betting. Since many bookmakers have options on predicting both the exact score and margin of victory of a match and if they do increase the betting success it gives them an added usage over predicting the toto probability directly.

References

Abramowitz, M. and I. A. Stegun (1965). *Handbook of mathematical functions, with formulas, graphs, and mathematical tables.* New York: Dover Publications.

Baio, G. and M. Blangiardo (2010). Bayesian hierarchical model for the prediction of football results. *Journal of Applied Statistics 37*(2), 253–264.

Buchdahl, J. (2003). *Fixed odds sports betting: Statistical forecasting and risk management.* Summersdale Publishers LTD-ROW.

Crowder, M., M. Dixon, A. Ledford, and M. Robinson (2002). Dynamic modelling and prediction of english football league matches for betting. *Journal of the Royal Statistical Society: Series D (The Statistician) 51*(2), 157–168.

Dixon, M. J. and S. G. Coles (1997). Modelling association football scores and inefficiencies in the football betting market. *Journal of the Royal Statistical Society: Series C (Applied Statistics) 46*(2), 265–280.

Elo, A. E. (1978). *The rating of chessplayers, past and present.* Arco Pub.

Fahrmeir, L. and G. Tutz (1994). Dynamic stochastic models for time-dependent ordered paired comparison systems. *Journal of the American Statistical Association 89*(428), 1438–1449.

Forrest, D. and R. Simmons (1999). *Making Up the Results: The Work of the Football Pools Panel, 1963-97.* Centre for the Study of Gambling and Commercial Gaming, University of Salford.

Goddard, J. and I. Asimakopoulos (2004). Forecasting football results and the efficiency of fixed-odds betting. *Journal of Forecasting 23*(1), 51–66.

Held, L. and R. Vollnhals (2005). Dynamic rating of european football teams. *IMA Journal of Management Mathematics 16*(2), 121–130.

Hvattum, L. M. and H. Arntzen (2010). Using ELO ratings for match result prediction in association football. *International Journal of forecasting 26*(3), 460–470.

Karlis, D. and I. Ntzoufras (2003). Analysis of sports data by using bivariate poisson models. *Journal of the Royal Statistical Society: Series D (The Statistician) 52*(3), 381–393.

Karlis, D. and I. Ntzoufras (2009, August). Bayesian modelling of football outcomes: using the skellam's distribution for the goal difference. *IMA Journal of Management Mathematics 20*(2), 133–145.

Knorr-Held, L. (2000, July). Dynamic rating of sports teams. *Journal of the Royal Statistical Society: Series D (The Statistician) 49*(2), 261–276.

Koning, R. H. (2000). Balance in competition in dutch soccer. *Journal of the Royal Statistical Society: Series D (The Statistician) 49*(3), 419–431.

Koopman, S. J. and R. Lit (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 178*(1), 167–186.

Koopman, S. J. and R. Lit (2019, April). Forecasting football match results in national league competitions using score-driven time series models. *International Journal of Forecasting 35*(2), 797–809.

Kruschke, J. K. (2011). *Doing Bayesian data analysis : a tutorial with R and BUGS.* Burlington, MA: Academic Press.

Maher, M. J. (1982). Modelling association football scores. *Statistica Neerlandica 36*(3), 109–118.

Manderson, A. A., K. Murray, and B. A. Turlach (2018, February). Dynamic bayesian forecasting of AFL match results using the skellam distribution. *Australian & New Zealand Journal of Statistics 60*(2), 174–187.

Owen, A. (2011, January). Dynamic bayesian forecasting models of football match outcomes with estimation of the evolution variance parameter. *IMA Journal of Management Mathematics 22*(2), 99–113.

Pollard, R., B. Benjamin, and C. Reep (1977). *Sport and the negative binomial distribution*, pp. 188–195. North Holland: Amsterdam.

Rue, H. and O. Salvesen (2000). Prediction and retrospective analysis of soccer matches in a league. *Journal of the Royal Statistical Society: Series D (The Statistician) 49*(3), 399–418.

Skellam, J. G. (1946). The frequency distribution of the difference between two poisson variates belonging to different populations. *Journal of the Royal Statistical Society. Series A 109 Pt 3*, 296.

Weisstein, E. W. (2002). *Modified Bessel function of the first kind.*

A Appendix

A.1 Data description

Table A.1, A.2 and A.3 report the descriptive statistics of the different seasons separately while Table A.4 shows the statistics of all seasons combined for the different observational variables home goals, away goals and the goal difference. The descriptive statistics in these tables are in order the minimum, maximum, median, mean, standard deviation, skewness and kurtosis. Comparing each of these statistics we can notice several things.

	Home Goals	Away Goals	Goal difference
min	0	0	-6
max	6	7	5
median	1	1	0
mean	1.5973	1.2026	0.3947
std.dev	1.3070	1.2279	1.9073
skewness	0.7353	1.1550	-0.2504
kurtosis	0.1656	1.4632	0.1532

Table A.1: Descriptive statistics of the 2016/2017 English Premier League season.

	Home Goals	Away Goals	Goal difference
min	0	0	-6
max	7	6	5
median	1	1	0
mean	1.5316	1.14737	0.3842
std.dev	1.3401	1.1778	1.8956
skewness	0.9530	1.0378	0.02808
kurtosis	0.6284	0.7108	0.3178

Table A.2: Descriptive statistics of the 2017-2018 English Premier League season.

	Home Goals	Away Goals	Goal difference
min	0	0	-5
max	6	6	6
median	1	1	0
mean	1.5684	1.2526	0.3158
std.dev	1.3128	1.1800	1.9152
skewness	0.8005	0.9640	-0.06852
kurtosis	0.3531	0.8818	0.02722

Table A.3: Descriptive statistics of the 2018-2019 English Premier League season.

Looking at the minimums and maximums we can see that if the maximum amount of Home Goals equals the maximum amount in Goal difference you can infer what the biggest home victory was that season. An example of this can be seen in Table A.3 where the maximums of both the Home Goals and Goal difference were 6 which infers that the biggest home win in the

	Home Goals	Away Goals	Goal difference
min	0	0	-6
max	7	7	6
median	1	1	0
mean	1.5658	1.2009	0.3649
std.dev	1.3192	1.1952	1.9047
skewness	0.8331	1.0581	-0.09814
kurtosis	0.3941	1.0577	0.1728

Table A.4: Descriptive statistics of the 2016-2017, 2017-2018 and 2018-2019 English Premier League seasons combined.

2018-2019 season is a 6-0 victory. The equivalent is true for the maximum amount of Away Goals and the minimum amount in Goal difference. An example of this can be seen in Table A.3 where the minimum Away Goals and minimum amount in Goal difference are 6 and -6 respectively, which would indicate that the biggest away victory had a score of 6-0. Note that if this is not the case this cannot be automatically inferred and in Table A.3 the maximum Away Goals is 6, which is one higher than the minimum Goal difference. One might believe that the biggest away victory might be 6-1, however if we go look into the data the result of the match with 6 away goals was actually 6-2 and the biggest away win was a 5-0 victory.

Besides inferring the biggest wins of the season, we can see from these statistics that in each season and over all seasons, the mean of the Home Goals is higher than that of the Away Goals and thus the mean of the Goal difference is also significantly above zero. The median, however is the same each season for Home Goals and Away Goals and the Goal difference is zero. This seems to suggest that although the home advantage is present it is not the only deciding factor for the result.

The standard deviation (std. dev) shows the volatility of the observational variable. As it turns out for all seasons separately and combined the volatility of Home Goals is higher than the Away Goals. Since both variables are unable to go below zero and the kurtosis from Home Goals is also higher than the kurtosis of Away Goals one can conclude that the high volatility is a further indication that the home advantage is a present phenomenon in the English Premier League. The overall higher volatility for the Goal difference than both other variables can be simply attributed to the fact that the Goal difference can be both positive or negative due to its definition.

The last two statistics skewness and kurtosis in all these tables are also interesting. Skewness seems to be slightly more positive for Away Goals than Home Goals which also suggests that the away teams seem to score less goals especially if you consider it in combination with the standard deviation. The skewness of Goal difference meanwhile is mostly close to zero, but it is slightly more negative in the 2016-2017 season suggesting that generally the distribution is quite symmetrical, but in the 2016-2017 season slightly more often on average the team playing at home scored more goals. For all variables over all seasons, the kurtosis is lower than the kurtosis of a normal distribution of 3. Meaning that they are all less peaked around their mean than the normal distribution and thus that it is more likely to differ from their means than if they were normally distributed which is consistent with what one would expect, because although given the various means and medians the average result would be a 1-1 draw this is usually not the most likely result of a single match.

Figure A.1 shows the goals scored and conceded of each team. By glancing at these graphs one can easily separate the top teams from the lower ranked teams. What also can be seen is that for some teams the amount of goals scored and the amount of goals conceded do not change much from season to season (for example Bournemouth), however for teams like Burnley, Crystal Palace, Leicester and West Ham it does change quite substantially which seems to encourage the use of the inter-season variance discussed in section 2.4. Note that in line with the handling of the promotion and relegation mentioned in section 3.1 some teams share their graph with another team, this team being the same team that they share their team number with. Interesting to see in these graphs is whether there is a big difference between the amount of goals scored and conceded as the graph starts to depict another team. This difference does seem quite substantial for some but generally not more than non-relegated teams generally change between seasons. An exception to this and where there is a biggest difference between teams is Wolves who both seem to score more goals and concede less than Swansea did in the 2 seasons before relegating. Although in the beginning of the season there are some narrow defeats for the Wolves. It will be interesting to see how the model handles this. But it seems that it can be expected that this would lead to the predictions for Wolves to not be quite as good as others might be. For other teams however if one looks at the graphs it looks as the same team. A good example of this is Cardiff that seem to have quite similar amounts of goals scored and conceded as Stoke had the season before. This should mean that the predictions of Cardiff should be relatively close to the actual results compared to predictions of teams that have not significantly changed between seasons.

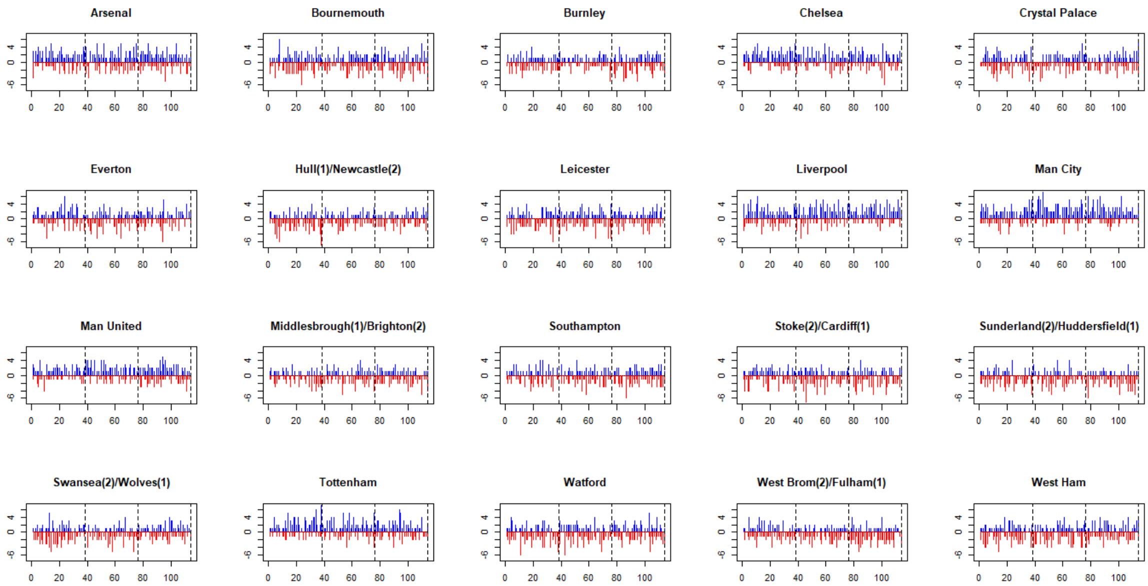

Figure A.1: Amount of goals scored (blue lines) and conceded red lines) over the 2016-2017, 2017-2018 and 2018-2019 English Premier League season by each team. Team1(x)/Team2(y) = Team1 played the first x of the 3 seasons in the Premier League and Team2 played the last y of the 3 seasons in the Premier League. Black vertical dashed lines indicate the end of a season.

A.2 Posterior means of Defence and Attacking strength

In this section, we present the posterior mean of the defense and attacking strengths of different teams across the 3 seasons.

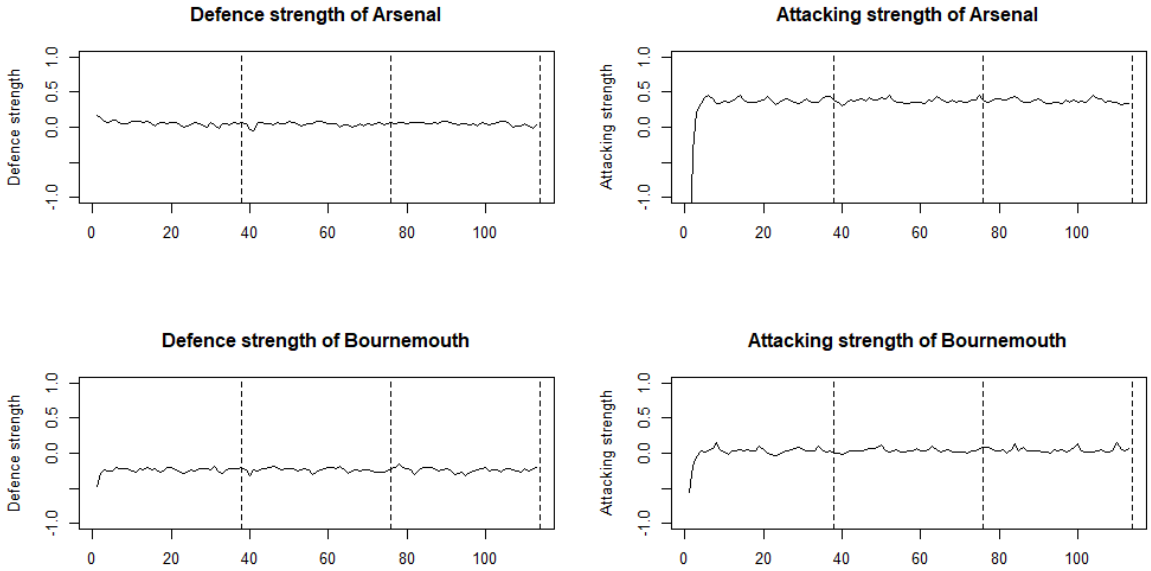

Figure A.2: Posterior means of the defense (left) and attacking (right) strength for Burnley (top) and Chelsea (bottom) for the 3 seasons.

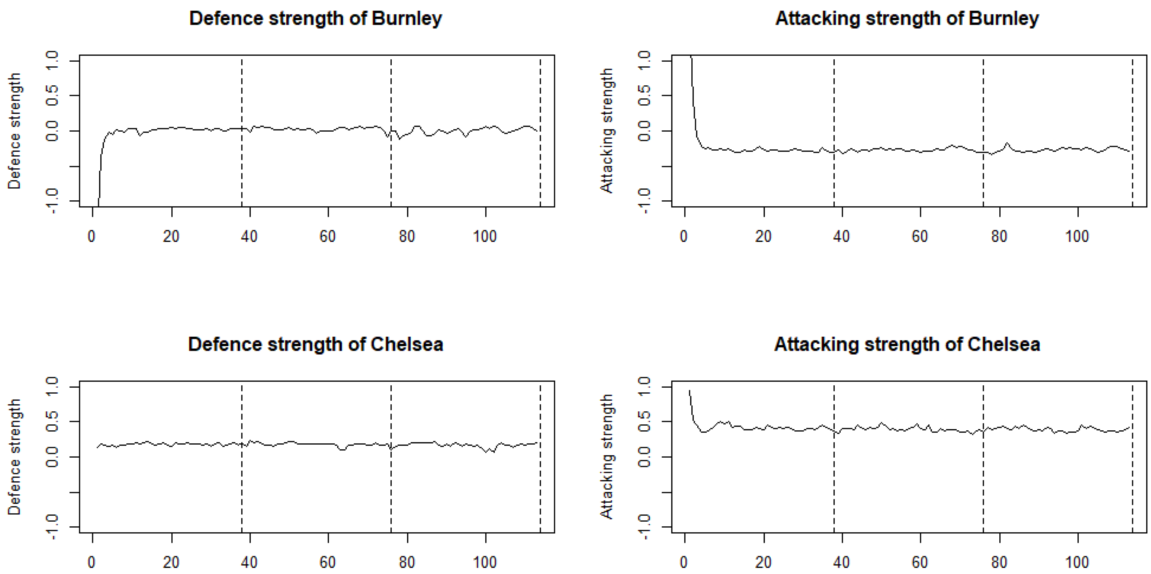

Figure A.3: Posterior means of the defense (left) and attacking (right) strength for Arsenal (top) and Bournemouth (bottom) for the 3 seasons.

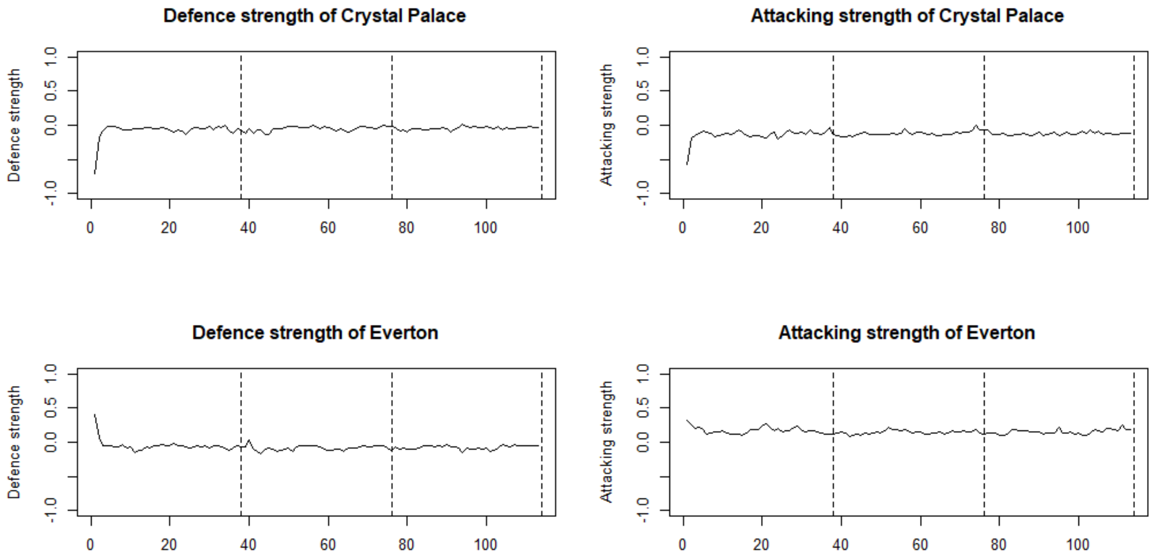

Figure A.4: Posterior means of the defense (left) and attacking (right) strength for Crystal Palace (top) and Everton (bottom) for the 3 seasons.

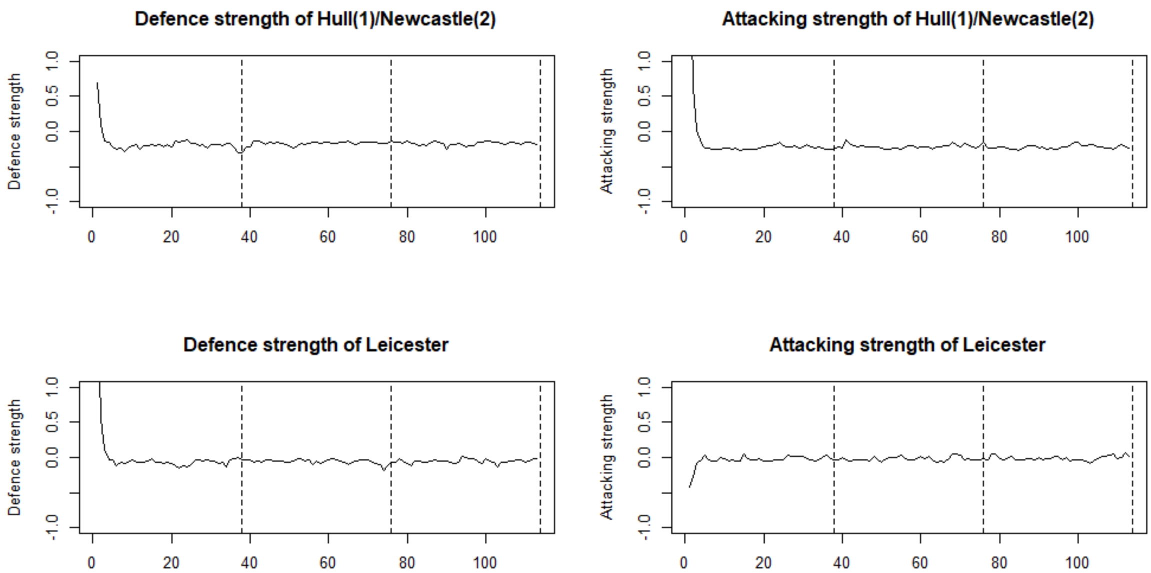

Figure A.5: Posterior means of the defense (left) and attacking (right) strength for Hull and Newcastle (top) and Leicester (bottom) for the 3 seasons.

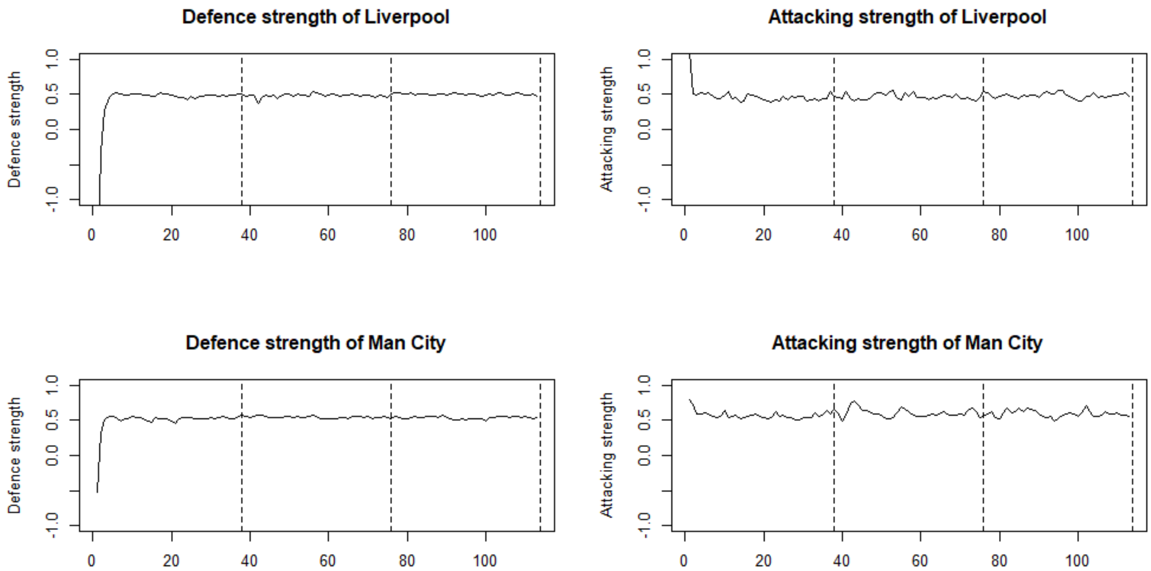

Figure A.6: Posterior means of the defense (left) and attacking (right) strength for Liverpool (top) and Manchester City (bottom) for the 3 seasons.

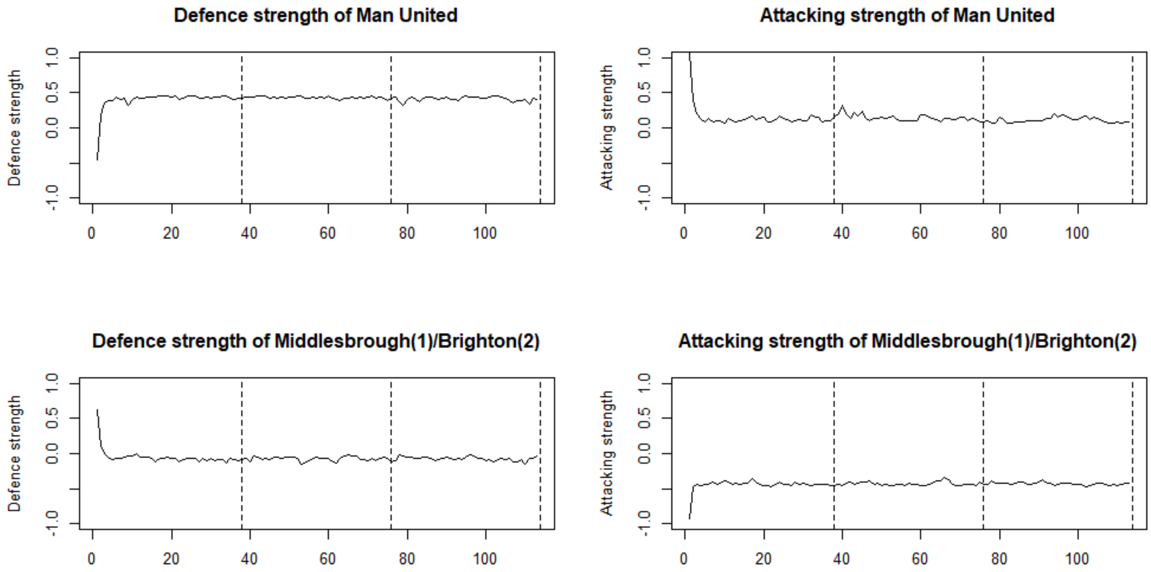

Figure A.7: Posterior means of the defense (left) and attacking (right) strength for Manchester United (top) and Middlesbrough and Brighton (bottom) for the 3 seasons.

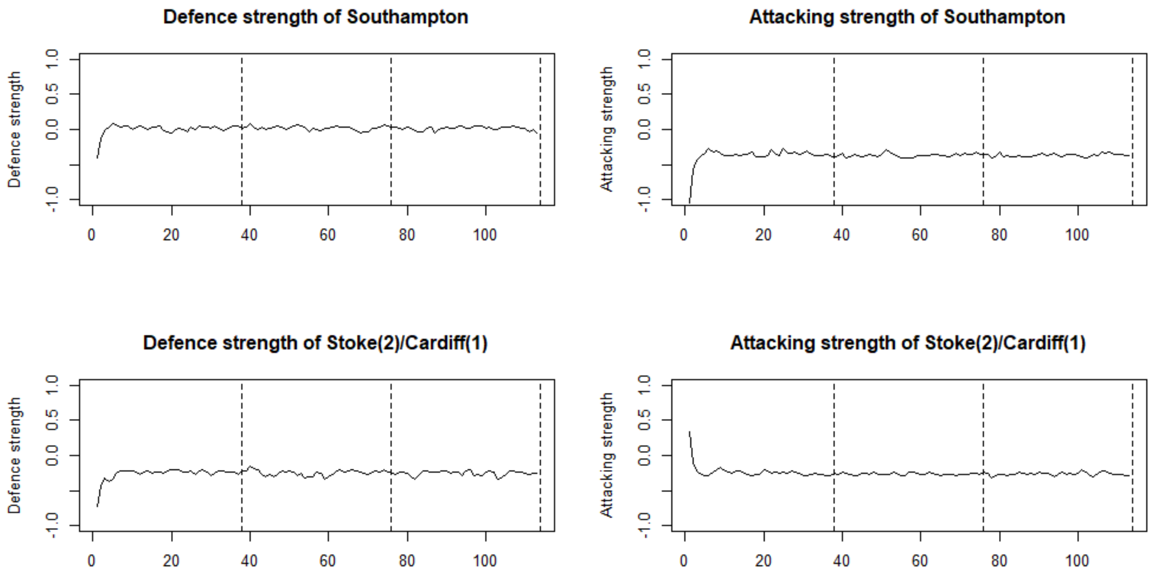

Figure A.8: Posterior means of the defense (left) and attacking (right) strength for Southampton (top) and Stoke and Cardiff (bottom) for the 3 seasons.

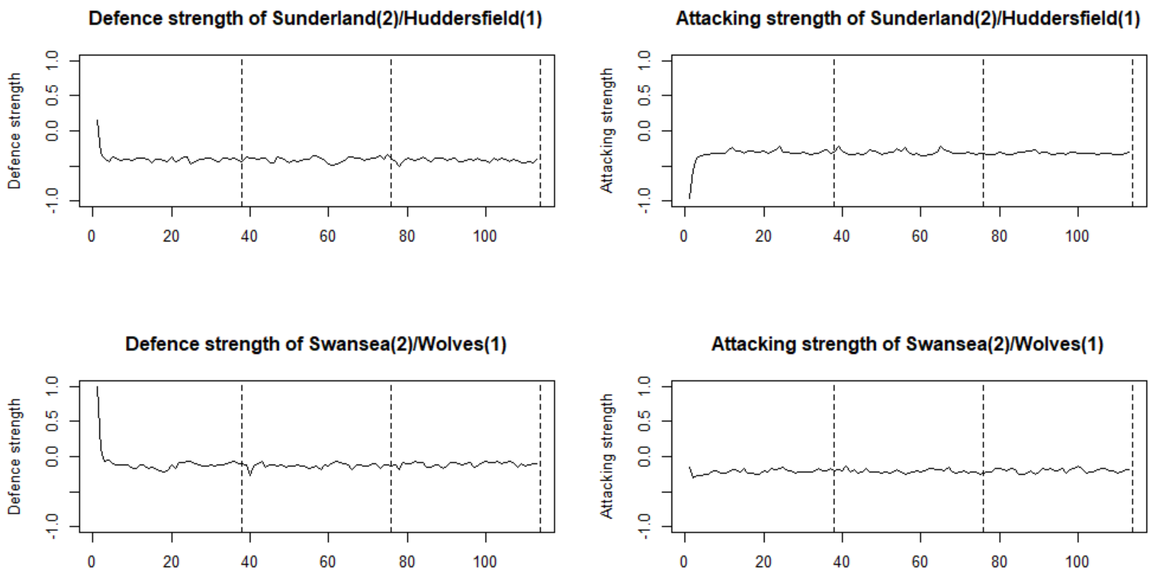

Figure A.9: Posterior means of the defense (left) and attacking (right) strength for Sunderland and Huddersfield (top) and Swansea and Wolves (bottom) for the 3 seasons.

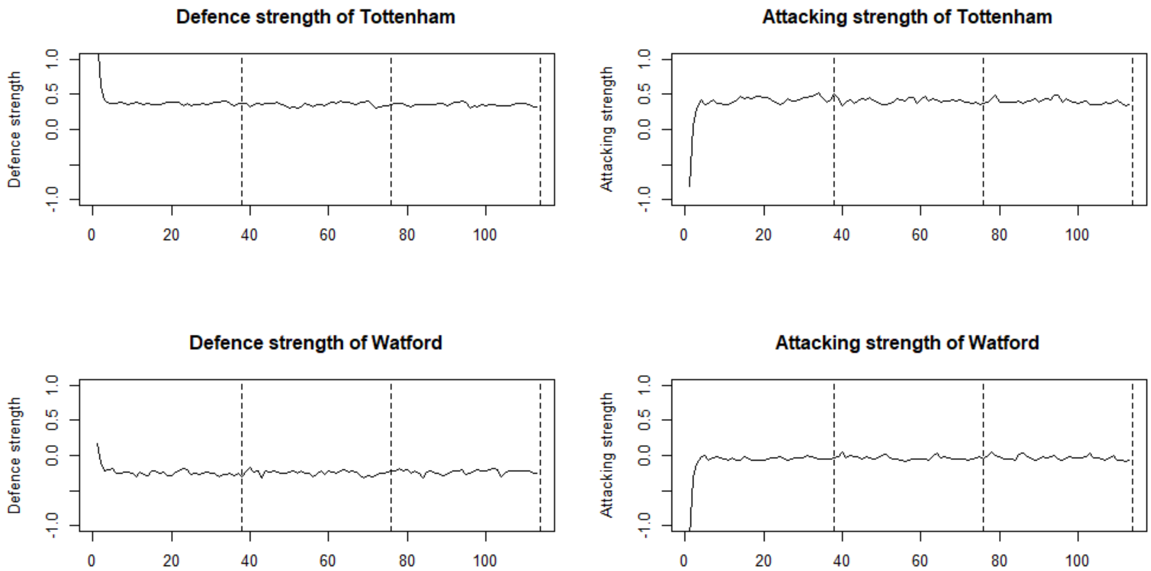

Figure A.10: Posterior means of the defense (left) and attacking (right) strength for Tottenham (top) and Watford (bottom) for the 3 seasons.

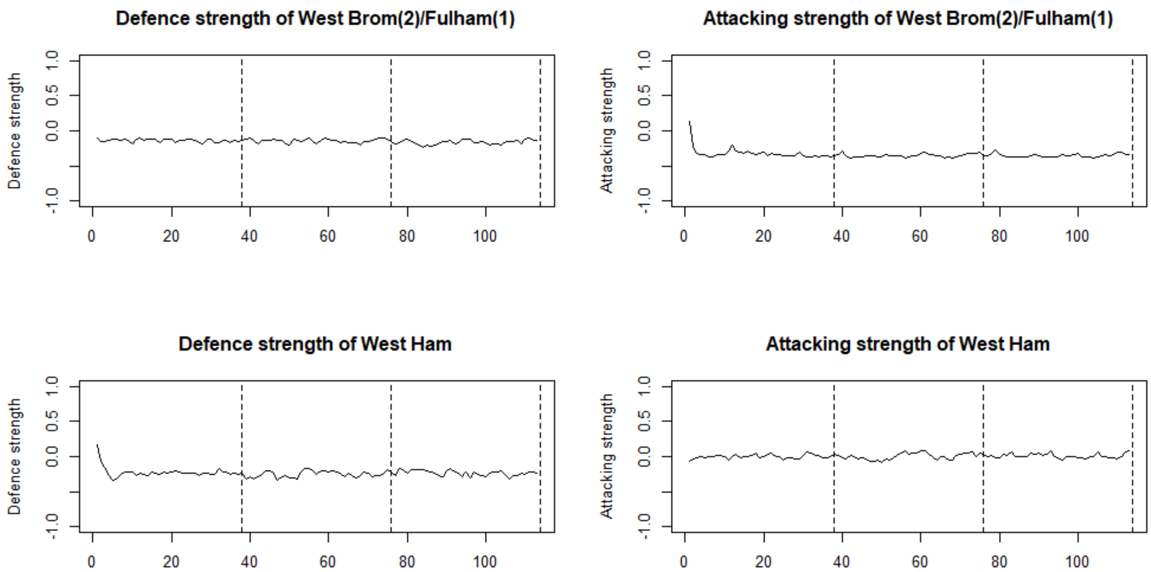

Figure A.11: Posterior means of the defense (left) and attacking (right) strength for West Brom and Fulham (top) and West Ham (bottom) for the 3 seasons.

A.3 Traceplot and density estimates of Goal Attack and Defend parameters

In this section, we present the traceplots and density estimates for the Goal Attack and Defend parameters for each of the top 5 teams in the final standings of the 2018-2019 English Premier League season at three different time points. Specifically those time points are the first round, the mid-season round (round 19) and the last round of the 2018-2019 season.

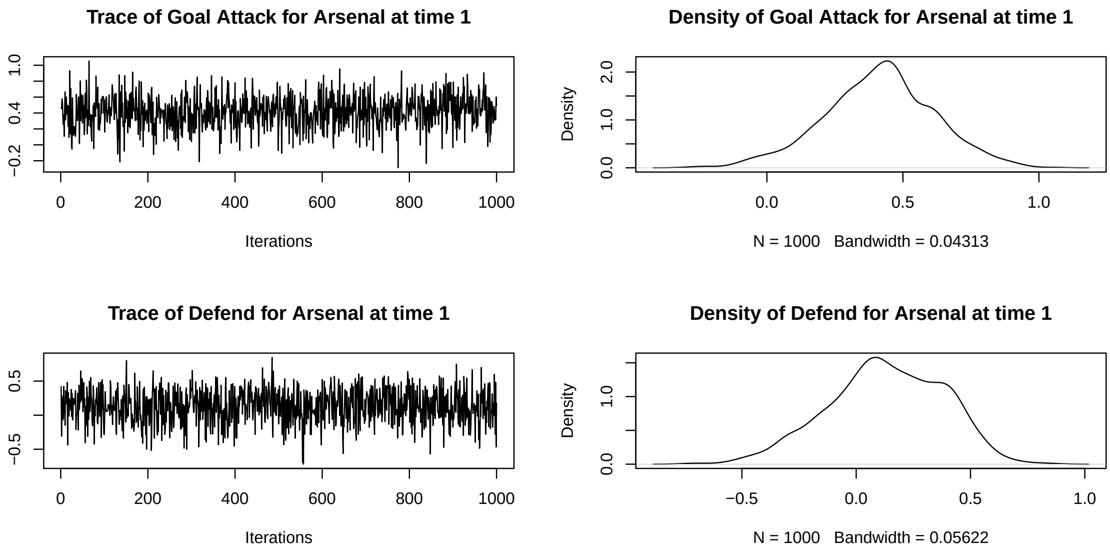

Figure A.12: Trace plot of the goal attack (top left) and of the defend (bottom left); density estimates of the goal attack (top right) and of the defend (bottom right) for Arsenal at the first round.

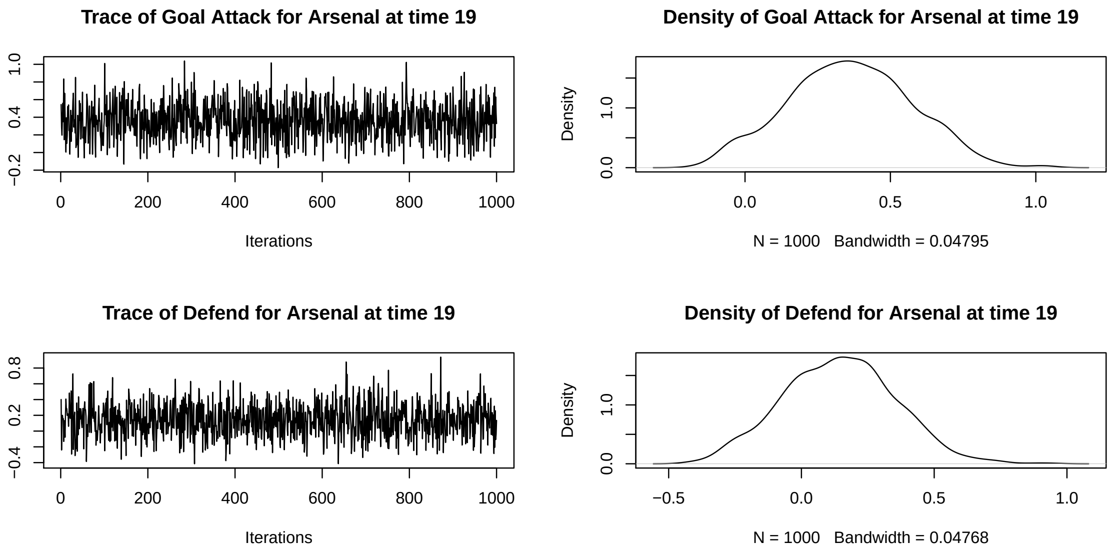

Figure A.13: Trace plot of the goal attack (top left) and of the defend (bottom left); density estimates of the goal attack (top right) and of the defend (bottom right) for Arsenal at the mid-season round.

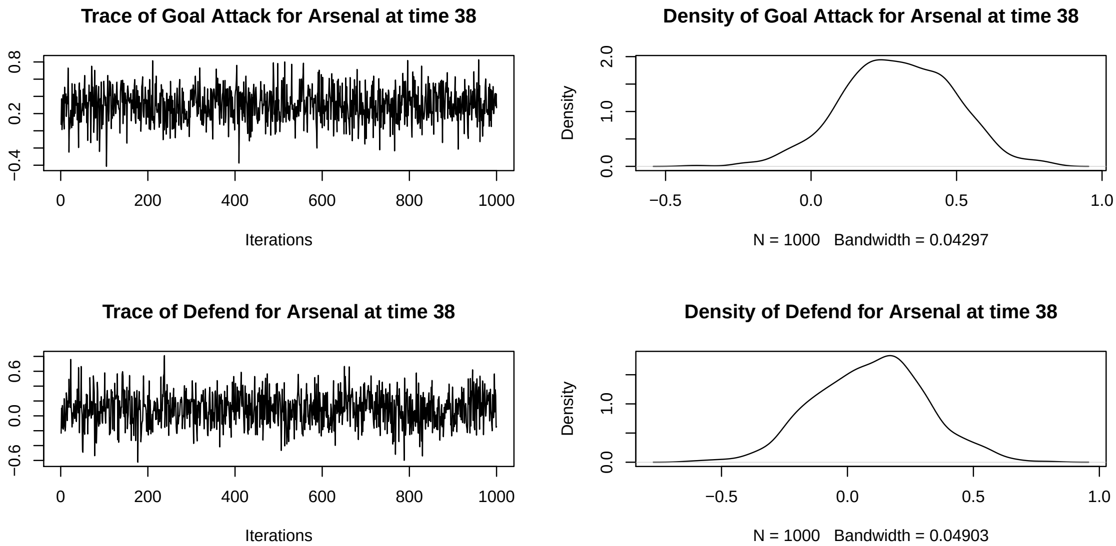

Figure A.14: Trace plot of the goal attack (top left) and of the defend (bottom left); density estimates of the goal attack (top right) and of the defend (bottom right) for Arsenal at the last round.

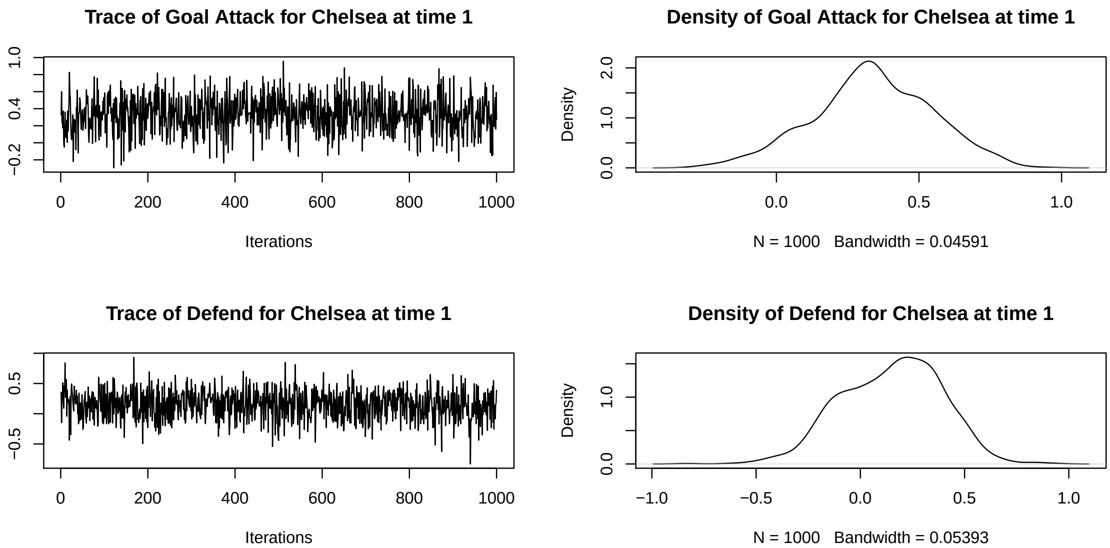

Figure A.15: Trace plot of the goal attack (top left) and of the defend (bottom left); density estimates of the goal attack (top right) and of the defend (bottom right) for Chelsea at the first round.

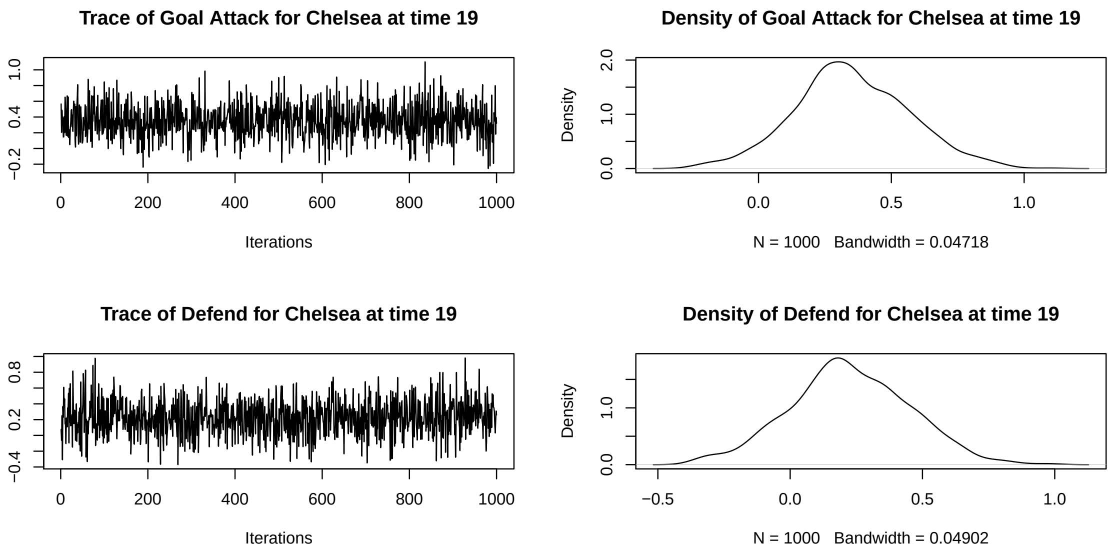

Figure A.16: Trace plot of the goal attack (top left) and of the defend (bottom left); density estimates of the goal attack (top right) and of the defend (bottom right) for Chelsea at the mid-season round.

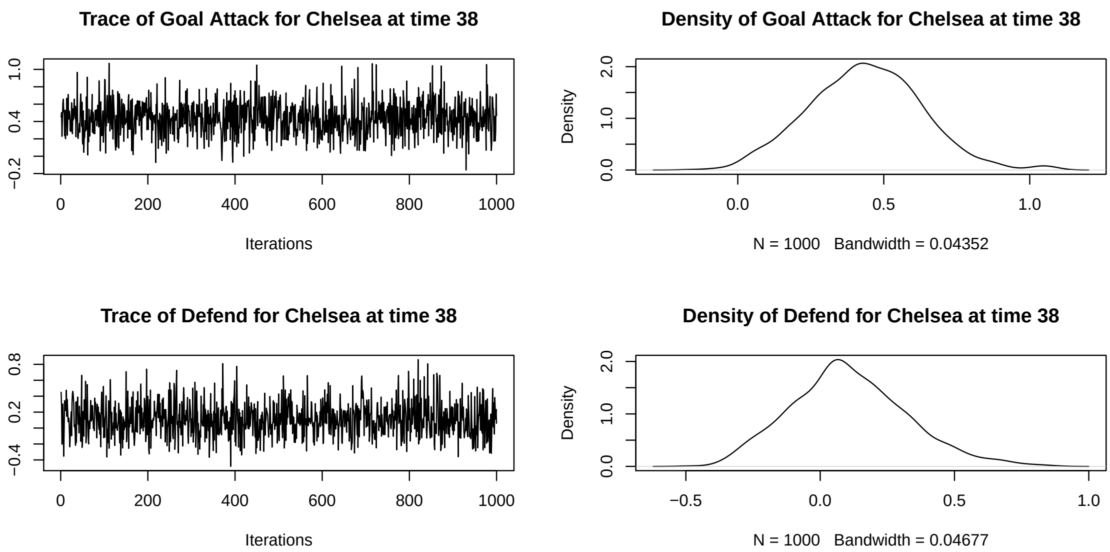

Figure A.17: Trace plot of the goal attack (top left) and of the defend (bottom left); density estimates of the goal attack (top right) and of the defend (bottom right) for Chelsea at the last round.

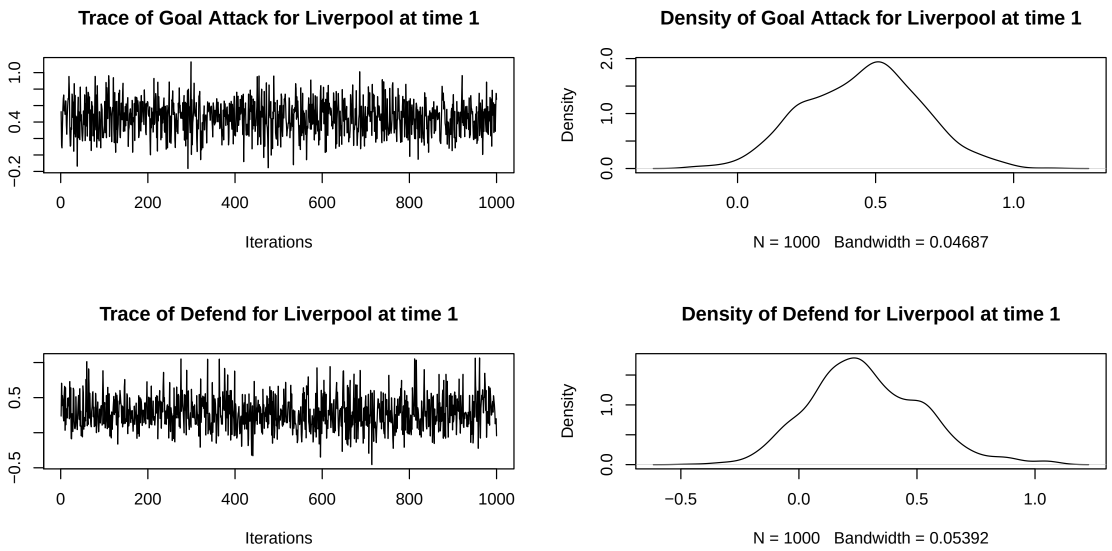

Figure A.18: Trace plot of the goal attack (top left) and of the defend (bottom left); density estimates of the goal attack (top right) and of the defend (bottom right) for Liverpool at the first round.

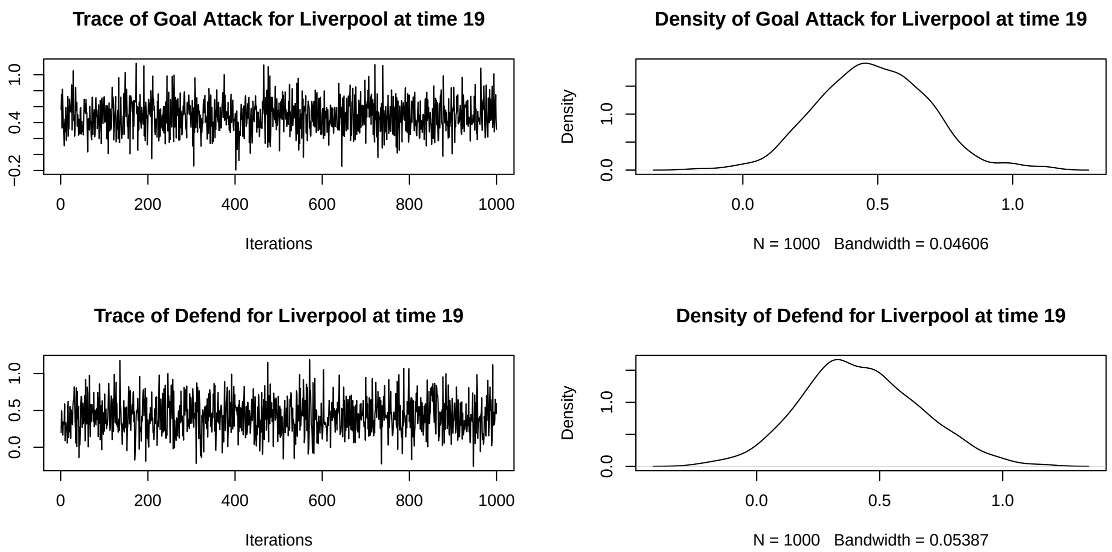

Figure A.19: Trace plot of the goal attack (top left) and of the defend (bottom left); density estimates of the goal attack (top right) and of the defend (bottom right) for Liverpool at the mid-season round.

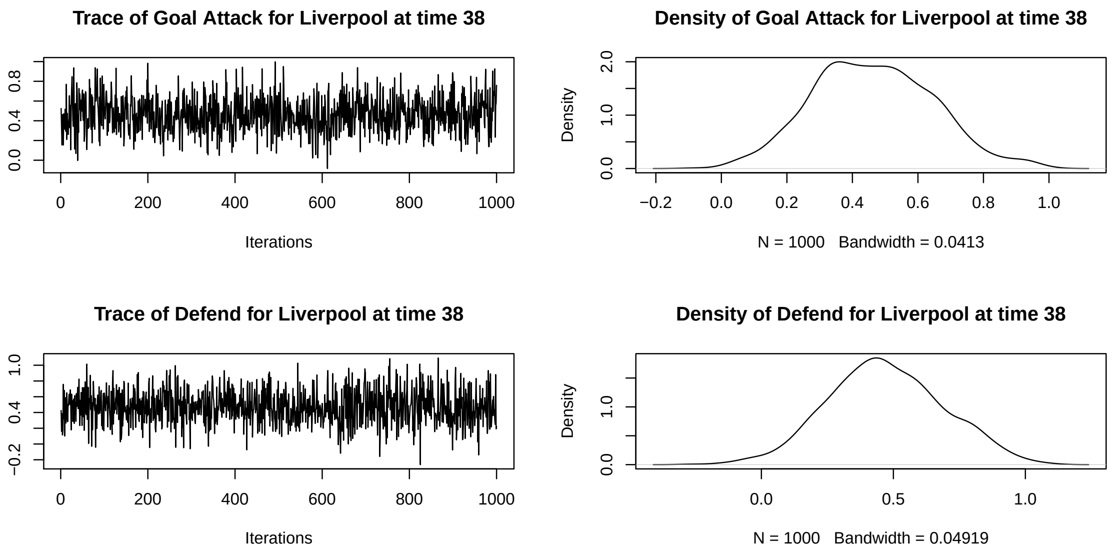

Figure A.20: Trace plot of the goal attack (top left) and of the defend (bottom left); density estimates of the goal attack (top right) and of the defend (bottom right) for Liverpool at the last round.

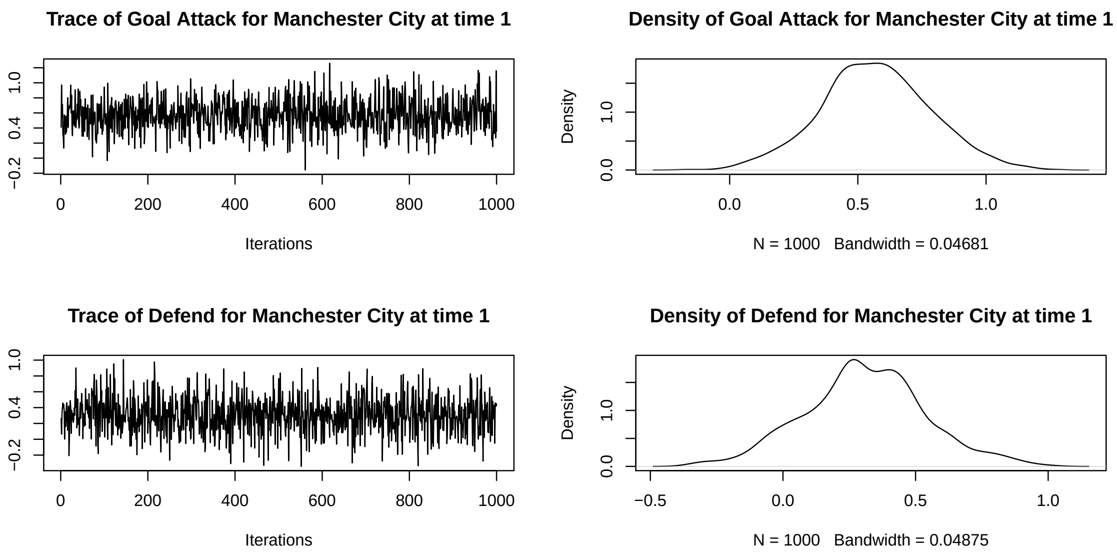

Figure A.21: Trace plot of the goal attack (top left) and of the defend (bottom left); density estimates of the goal attack (top right) and of the defend (bottom right) for Manchester City at the first round.

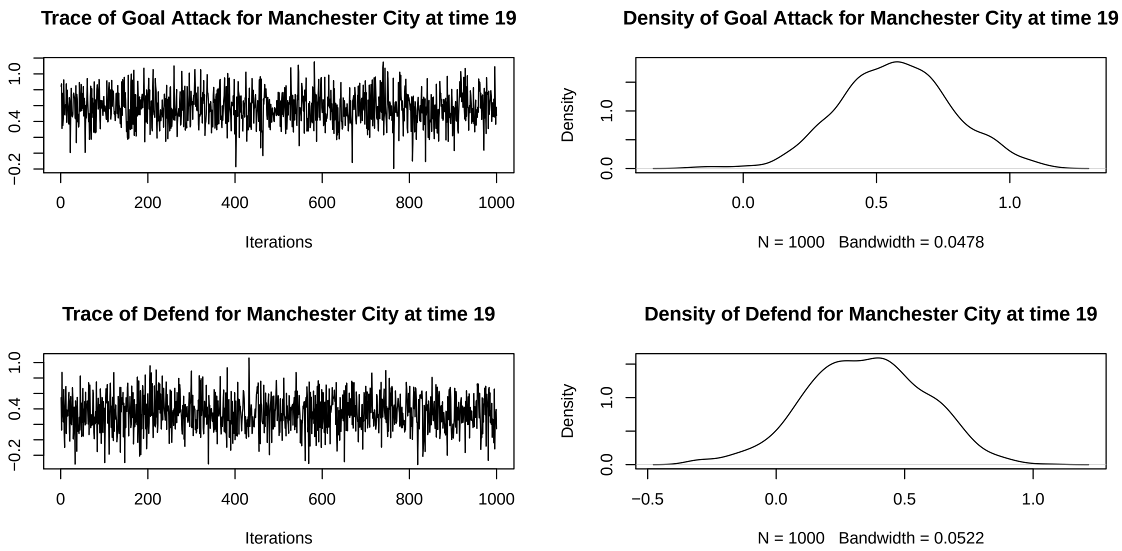

Figure A.22: Trace plot of the goal attack (top left) and of the defend (bottom left); density estimates of the goal attack (top right) and of the defend (bottom right) for Manchester City at the mid-season round.

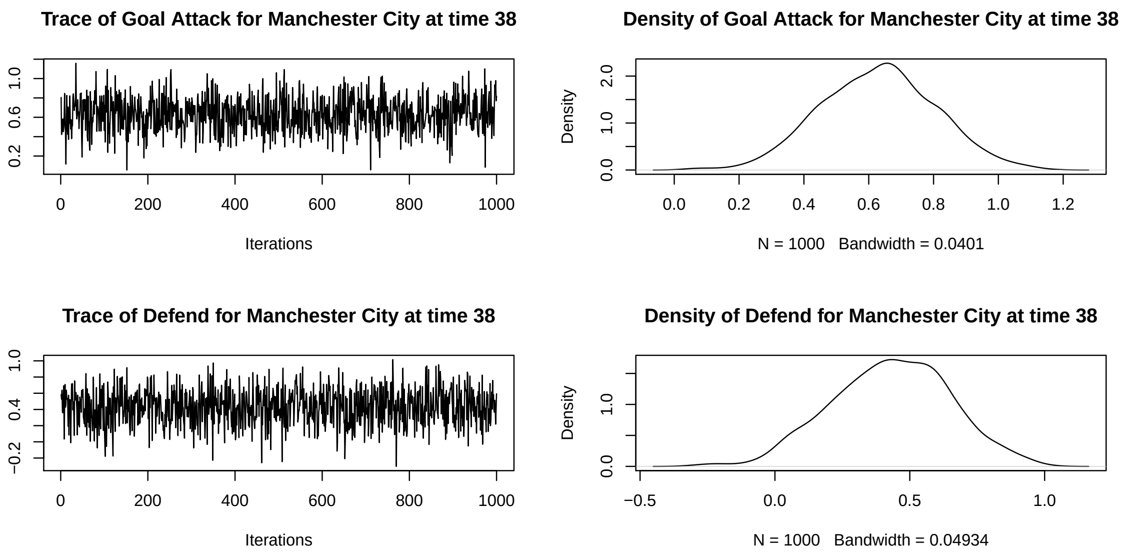

Figure A.23: Trace plot of the goal attack (top left) and of the defend (bottom left); density estimates of the goal attack (top right) and of the defend (bottom right) for Manchester City at the last round.

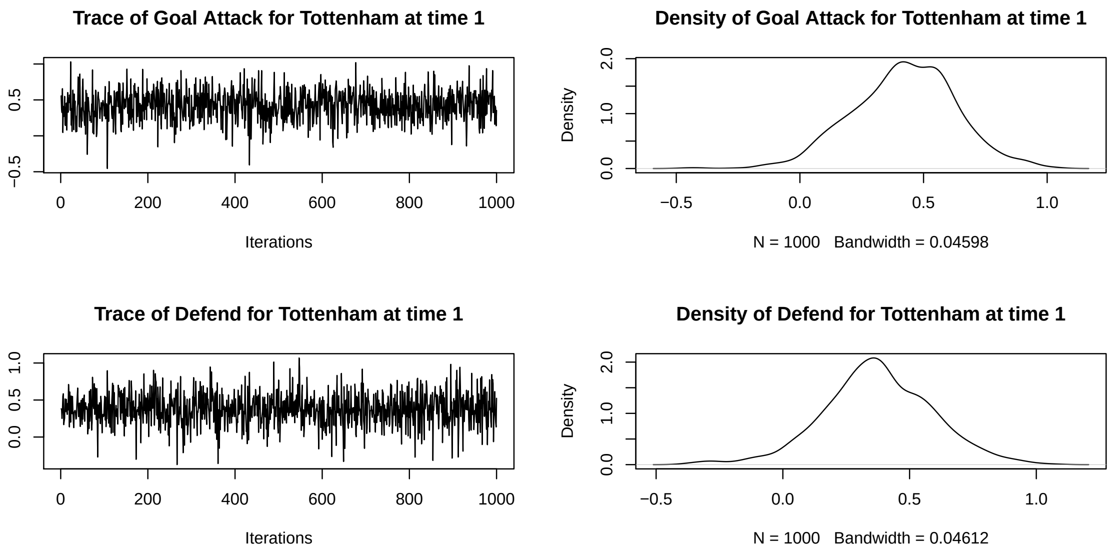

Figure A.24: Trace plot of the goal attack (top left) and of the defend (bottom left); density estimates of the goal attack (top right) and of the defend (bottom right) for Tottenham at the first round.

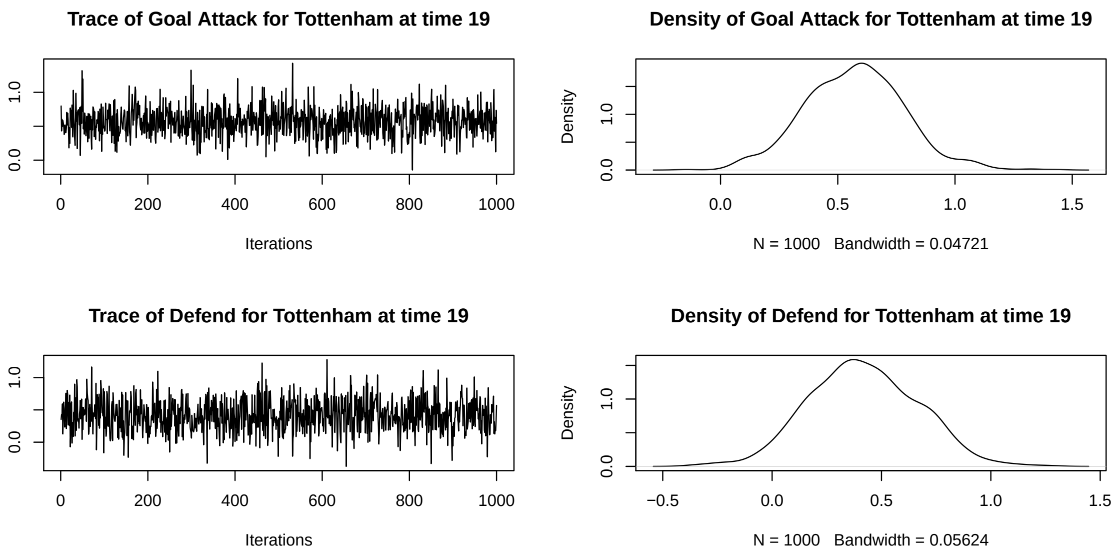

Figure A.25: Trace plot of the goal attack (top left) and of the defend (bottom left); density estimates of the goal attack (top right) and of the defend (bottom right) for Tottenham at the mid-season round.

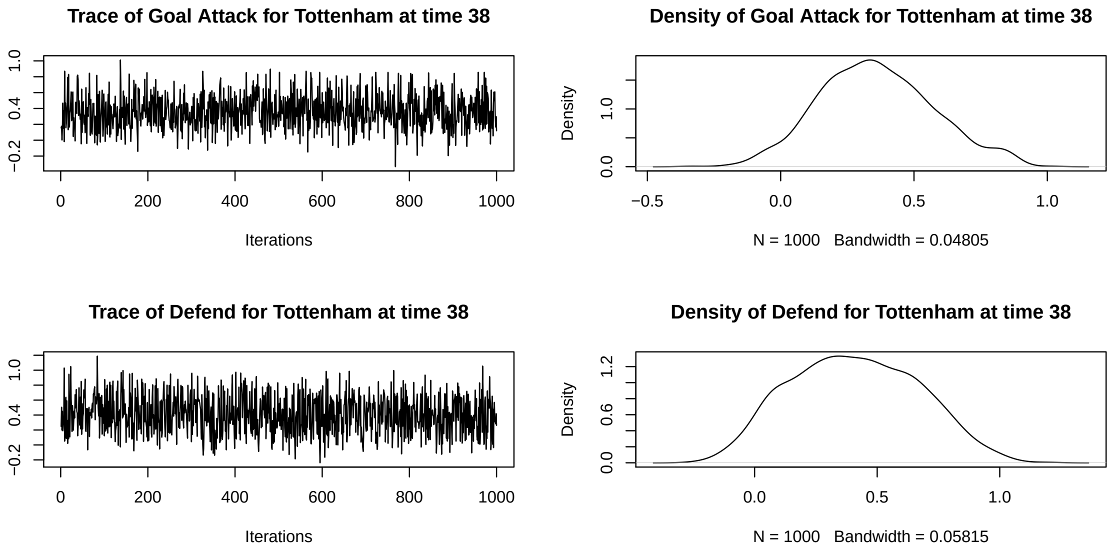

Figure A.26: Trace plot of the goal attack (top left) and of the defend (bottom left); density estimates of the goal attack (top right) and of the defend (bottom right) for Tottenham at the last round.