# Predicting Premier League Football Results Over the Last Decade.

Isaac Scott
32120255

MSc Statistics

Lancaster University
United Kingdom

7th September 2018

# Contents

# 1 Introduction

Although the market for online sports betting has experienced consistent growth over the course of the past decade or so, game prediction, particularly for the sport of football, is the subject of surprisingly few pieces of literature. This in part is due to the complexity of the subject, and the fact that software to model and describe the issue has only recently become sophisticated enough to be sufficiently accurate.

Initial, more unsophisticated attempts at modelling the number of goals scored in a football match, such as in [10] Moroney (1951) refuted the postulation of utilising Poisson models for home and away goals in favour of what they termed a "modified Poisson" distribution, alternatively named the Negative Binomial distribution.
[11] Reep, Pollard and Benjamin (1971) developed the idea of employing Negative Binomial models further, applying said models to the subject of home and away goals scored within four English Football League First Division seasons, as well as other subjects. These included successful pass attempts within association football games, and modelling other ball games. Both [10] Moroney (1951) and [11] Reep, Pollard and Benjamin (1971) found the Negative Binomial models to be a close fit for the counts of home and away goals, subsequently drawing the conclusion that the same Negative Binomial may be employed to model the number of goals scored by each team, regardless of the team's ability.

This objective model, which does not account for individual team abilities came under scrutiny though, in the seminal paper by [9] Maher (1982). Within Maher's paper, the Negative Binomial models posited by previous papers were rejected in favour of independent Poisson models for the count data of home and away goals, facilitating the production of attack and defence strength scores for each team under study (Maher utilised the top four leagues of English football for three seasons, 1973-1975).
Maher initially attempted employing both Attack and Defence Strength parameters for both home and away teams, which resulted in an 88-parameter model in a league of 22 teams (4 parameters per team, namely home and away attack and defence strengths) reduced to 86 by constraints. However, this model was found to be sub-optimal due to its complexity when Chi-squared ($\chi^2$) tests were employed to hierarchal subsets of its parameters. Thus, Maher instead favoured a model only including singular "Attack Strength" and "Defence Strength" parameters for each team.

As well as utilising "Attack Strength" and "Defence Strength" parameters for each team, a common "Home Ground Advantage" parameter was used within the model, reflecting the apparent increase in a team's performance when playing at home. This parameter was found to be significant by Maher, but common to all teams, as opposed to the suggestions of [3] Norman and Clarke (1995), in which "Home Advantage" scores were found to vary between teams within a non-parametric format.

As well as parametrising the prior univariate Poisson models in the way mentioned previously, Maher also investigated possible breaches of the independence assumption between

home and away goal counts. This issue was further explored by [5] Dixon & Coles (1997), as well as [7] Karlis & Ntzoufras (2003), who pointed out that "In team sports, such as football and water-polo, it is reasonable to assume two outcome variables are correlated since the two teams interact."
These papers dealt with the issue of potential correlation via the use of Bivariate Poisson distributions, with Maher finding models to perform better than their Univariate Poisson counterparts when a correlation parameter of around 0.2 is used.
Dixon & Coles, as well as Karlis & Ntzoufras, paid particular attention to low-scoring games, which were found to be significantly more correlated than when considering the scores in every game. Dixon & Coles included a score-specific correlation function to account for correlation within 0-0, 0-1, 1-0 and 1-1 games. Karlis & Ntzoufras simply included an extra correlation parameter to account for this, as well as "inflating" the Poisson models to account for the under-estimation of draws.

Within this paper, we aim to compare and contrast the models mentioned within this section, fully investigating their validity, as well as variations of the models to find the optimal model in describing the results of English Premier League football matches over the last 9 seasons. As well as analysing pre-existing models, we hope also to investigate the utility of a Bayesian approach to the same task. Bayesian approaches to game predicton in particular have had little-to-no coverage within literature to our knowledge, despite their utilisation of prior knowledge seeming well-suited to the task of prediction within the sport of football. We hope to explore both the utility of non-sequential (using MCMC methods) and sequential forms of Bayesian modelling in the context of Premier League football, and observe how well these models may be used to predict compared to the estimates of modern-day betting companies.

# 2 Data

Within our study dataset, we are provided with data from nine English Premier League seasons, from the outset of the 2009/10 season to the conclusion of the 2017/18 season.
For each season, the data provides the result of each match within each of the 38 rounds of play. These results include the number of home and away goals (HG, AG) scored during each game, subsequently providing the difference in score (HG-AG) for each game. These measures are displayed within Figure 1 below for the $8^{th}$ season of study, 2016/17.

These goal counts can be used simply to determine the winning team in each matchup, as shown within Figure 2, where an entry of 1 signifies a home win, 0 a draw and -1 an away win.
These measures are to be used later in the construction of both frequentist and Bayesian parametric models for the data.

| | Arsenal | Bournemouth | Burnley | Chelsea | Crystal Palace | Everton | Hull | Leicester | Liverpool | Man City | Man United | Middlesbrough | Southampton | Stoke | Sunderland | Swansea | Tottenham | Watford | West Brom | West Ham | Total |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| *Arsenal* | | 2 | 1 | 3 | 2 | 2 | 2 | 1 | -1 | 0 | 2 | 0 | 1 | 2 | 2 | 1 | 0 | -1 | 1 | 3 | 23 |
| *Bournemouth* | 0 | | 1 | -2 | -2 | 1 | 5 | 1 | 1 | -2 | -2 | 4 | -2 | 0 | -1 | 2 | 0 | 0 | 1 | 1 | 6 |
| *Burnley* | -1 | 1 | | 0 | 1 | 1 | 0 | 1 | 2 | -1 | -2 | 1 | 1 | 1 | 3 | -1 | -2 | 2 | 0 | -1 | 6 |
| *Chelsea* | 2 | 3 | 3 | | -1 | 5 | 2 | 3 | -1 | 1 | 4 | 3 | 2 | 2 | 4 | 2 | 1 | 1 | 1 | 1 | 38 |
| *Crystal Palace* | 3 | 0 | -2 | -1 | | -1 | 4 | 0 | -2 | -1 | -1 | 1 | 3 | 3 | -4 | -1 | -1 | 1 | -1 | -1 | -1 |
| *Everton* | 1 | 3 | 2 | -3 | 0 | | 4 | 2 | -1 | 4 | 0 | 2 | 3 | 1 | 2 | 0 | 0 | 1 | 3 | 2 | 26 |
| *Hull* | -3 | 2 | 0 | -2 | 0 | 0 | | 1 | 2 | -3 | -1 | 2 | 1 | -2 | -2 | 1 | -6 | 2 | 0 | 1 | -7 |
| *Leicester* | 0 | 0 | 3 | -3 | 2 | -2 | 2 | | 2 | 2 | -3 | 0 | 0 | 2 | 2 | 1 | -5 | 3 | -1 | 1 | 6 |
| *Liverpool* | 2 | 0 | 1 | 0 | -1 | 2 | 4 | 3 | | 1 | 0 | 3 | 0 | 3 | 2 | -1 | 2 | 5 | 1 | 0 | 27 |
| *Man City* | 1 | 4 | 1 | -2 | 5 | 0 | 2 | 1 | 0 | | 0 | 0 | 0 | 0 | 1 | 1 | 0 | 2 | 2 | 2 | 20 |
| *Man United* | 0 | 0 | 0 | 2 | 2 | 0 | 0 | 3 | 0 | -1 | | 1 | 2 | 0 | 2 | 0 | 1 | 2 | 0 | 0 | 14 |
| *Middlesbrough* | -1 | 2 | 0 | -1 | -1 | 0 | 1 | 0 | -3 | 0 | -2 | | -1 | 0 | 1 | 3 | -1 | -1 | 0 | -2 | -6 |
| *Southampton* | -2 | 0 | 2 | -2 | 2 | 1 | 0 | 3 | 0 | -3 | 0 | 1 | | -1 | 0 | 1 | -3 | 0 | -1 | -2 | -4 |
| *Stoke* | -3 | -1 | 2 | -1 | 1 | 0 | 2 | 0 | -1 | -3 | 0 | 2 | 0 | | 2 | 2 | -4 | 2 | 0 | 0 | 0 |
| *Sunderland* | -3 | -1 | 0 | -1 | -1 | -3 | 3 | 1 | 0 | -2 | -3 | -1 | -4 | -2 | | -2 | 0 | 1 | 0 | 0 | -18 |
| *Swansea* | -4 | -3 | 1 | 0 | 1 | 1 | -2 | 2 | -1 | -2 | -2 | 0 | 1 | 2 | 3 | | -2 | 0 | 1 | -3 | -7 |
| *Tottenham* | 2 | 4 | 1 | 2 | 1 | 1 | 3 | 0 | 0 | 2 | 1 | 1 | 1 | 4 | 1 | 5 | | 4 | 4 | 1 | 38 |
| *Watford* | -2 | 0 | 1 | -1 | 0 | 1 | 1 | 1 | -1 | -5 | 2 | 0 | -1 | -1 | 1 | 1 | -3 | | 2 | 0 | -4 |
| *West Brom* | 2 | 1 | 4 | -1 | -2 | -1 | 2 | -1 | -1 | -4 | -2 | 0 | -1 | 1 | 2 | 2 | 0 | 2 | | 2 | 5 |
| *West Ham* | -4 | 1 | 1 | -1 | 3 | 0 | 1 | -1 | -4 | -4 | -2 | 0 | -3 | 0 | 1 | 1 | 1 | -2 | 0 | | -12 |
| *Total* | 10 | -18 | -22 | 14 | -12 | -8 | -36 | -21 | 9 | 21 | 11 | -20 | -3 | -15 | -22 | -18 | 22 | -24 | -13 | -5 | |

Figure 1: Grid of Differences (HG-AG) for each game in the 2016/17 season. Home Team (left), Away Team (top)

| | Arsenal | Bournemouth | Burnley | Chelsea | Crystal Palace | Everton | Hull | Leicester | Liverpool | Man City | Man United | Middlesbrough | Southampton | Stoke | Sunderland | Swansea | Tottenham | Watford | West Brom | West Ham | Total wins |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| *Arsenal* | | 1 | 1 | 1 | 1 | 1 | 1 | 1 | -1 | 0 | 1 | 0 | 1 | 1 | 1 | 1 | 0 | -1 | 1 | 1 | 14 |
| *Bournemouth* | 0 | | 1 | -1 | -1 | 1 | 1 | 1 | 1 | -1 | -1 | 1 | -1 | 0 | -1 | 1 | 0 | 0 | 1 | 1 | 9 |
| *Burnley* | -1 | 1 | | 0 | 1 | 1 | 0 | 1 | 1 | -1 | -1 | 1 | 1 | 1 | 1 | -1 | -1 | 1 | 0 | -1 | 10 |
| *Chelsea* | 1 | 1 | 1 | | -1 | 1 | 1 | 1 | -1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 17 |
| *Crystal Palace* | 1 | 0 | -1 | -1 | | -1 | 1 | 0 | -1 | -1 | -1 | 1 | 1 | 1 | -1 | -1 | -1 | 1 | -1 | -1 | 6 |
| *Everton* | 1 | 1 | 1 | -1 | 0 | | 1 | 1 | -1 | 1 | 0 | 1 | 1 | 1 | 1 | 0 | 0 | 1 | 1 | 1 | 13 |
| *Hull* | -1 | 1 | 0 | -1 | 0 | 0 | | 1 | 1 | -1 | -1 | 1 | 1 | -1 | -1 | 1 | -1 | 1 | 0 | 1 | 8 |
| *Leicester* | 0 | 0 | 1 | -1 | 1 | -1 | 1 | | 1 | 1 | -1 | 0 | 0 | 1 | 1 | 1 | -1 | 1 | -1 | 1 | 10 |
| *Liverpool* | 1 | 0 | 1 | 0 | -1 | 1 | 1 | 1 | | 1 | 0 | 1 | 0 | 1 | 1 | -1 | 1 | 1 | 1 | 0 | 12 |
| *Man City* | 1 | 1 | 1 | -1 | 1 | 0 | 1 | 1 | 0 | | 0 | 0 | 0 | 0 | 1 | 1 | 0 | 1 | 1 | 1 | 11 |
| *Man United* | 0 | 0 | 0 | 1 | 1 | 0 | 0 | 1 | 0 | -1 | | 1 | 1 | 0 | 1 | 0 | 1 | 1 | 0 | 0 | 8 |
| *Middlesbrough* | -1 | 1 | 0 | -1 | -1 | 0 | 1 | 0 | -1 | 0 | -1 | | -1 | 0 | 1 | 1 | -1 | -1 | 0 | -1 | 4 |
| *Southampton* | -1 | 0 | 1 | -1 | 1 | 1 | 0 | 1 | 0 | -1 | 0 | 1 | | -1 | 0 | 1 | -1 | 0 | -1 | -1 | 6 |
| *Stoke* | -1 | -1 | 1 | -1 | 1 | 0 | 1 | 0 | -1 | -1 | 0 | 1 | 0 | | 1 | 1 | -1 | 1 | 0 | 0 | 7 |
| *Sunderland* | -1 | -1 | 0 | -1 | -1 | -1 | 1 | 1 | 0 | -1 | -1 | -1 | -1 | -1 | | -1 | 0 | 1 | 0 | 0 | 3 |
| *Swansea* | -1 | -1 | 1 | 0 | 1 | 1 | -1 | 1 | -1 | -1 | -1 | 0 | 1 | 1 | 1 | | -1 | 0 | 1 | -1 | 8 |
| *Tottenham* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | | 1 | 1 | 1 | 17 |
| *Watford* | -1 | 0 | 1 | -1 | 0 | 1 | 1 | 1 | -1 | -1 | 1 | 0 | -1 | -1 | 1 | 1 | -1 | | 1 | 0 | 8 |
| *West Brom* | 1 | 1 | 1 | -1 | -1 | -1 | 1 | -1 | -1 | -1 | -1 | 0 | -1 | 1 | 1 | 1 | 0 | 1 | | 1 | 9 |
| *West Ham* | -1 | 1 | 1 | -1 | 1 | 0 | 1 | -1 | -1 | -1 | -1 | 0 | -1 | 0 | 1 | 1 | 1 | -1 | 0 | | 7 |
| *Total wins* | 9 | 3 | 1 | 13 | 6 | 4 | 1 | 2 | 10 | 12 | 10 | 1 | 6 | 4 | 3 | 4 | 9 | 3 | 3 | 5 | |

Figure 2: Grid of Results (HG-AG) for each game in the 2016/17 season. Home Win = 1, Away Win = -1, Draw = 0.

# 3 Poisson Generalised Linear Models

## 3.1 Generalised Linear Models

Instead of beginning with the non-parametric Norman and Clarke model (as discussed within S1) we initially opt to model the seasonal data by utilising univariate Poisson Generalised Linear Models (GLMs).
A GLM is defined by its underpinning assumptions, which are as follows:

1. **Nature of Observations**: Observations are taken on a one-dimensional response variable, $Y$, indexed by $i$, with $i = 1,...,n$; with $p < n$ explanatory variables; $x_{1i}, ..., x_{pi}$.
2. **Independence of Response variables**: The responses within the model ($Y_i$, $i = 1,...,n$) are realisations of random variables, observed independently of each other.
3. **Exponential Family**: The conditional distribution of the response variable, $Y$ within the model belongs to the exponential family of distributions, with mean $\mu$ and scale pa-

rameter $\phi$. A distribution function belonging to the exponential family of distributions can be expressed in the form;

$$f(y|\theta, \phi) = \exp\left[\frac{y\theta - k(\theta)}{\phi} + c(y, \phi)\right] \tag{1}$$

In which $\theta$ and $\phi$ are parameters, and $k$ and $c$ are functions which ensure the distribution is proper.

4. **Linear Predictor**: A single linear function (the linear predictor, $\psi$) exists consisting of explanatory variables, through which these explanatory variables act upon $Y$. This is of the form;

$$\psi = \beta_1 x_{i1} + ... + \beta_p x_{ip} \tag{2}$$

5. **Link Function**: The mean $\mu$ of $Y$ and linear predictor $\psi$ are related via a smooth, invertible link function, $\boldsymbol{g}(\cdot)$ which satisfies $\boldsymbol{g}(\mu) = \psi$

## 3.2 Univariate Poisson Models

Modelling the outcome of each result in a particular season is initially done via the fitting of three differing models using a range of covariates in relation to the data.
Throughout this section, $i$ is used to signify the home team, with $1 \leq i \leq 20$. We use $j$ as a signifier for the corresponding away team, with $1 \leq j \leq 20$.

### 3.2.1 Model 1

The first of our univariate Poisson models utilises both Attacking Strength (AS) and Defensive Strength (DS) parameters for each team in conjunction with a common Home Ground Advantage parameter (HGA) shared by each team.
We use these covariates to model the number of home goals in a game between home team $i$ and away team $j$, $X_{ij}$, and the number of away goals in a match between teams $i$ and $j$, $Y_{ij}$.
We construct the model with the following formulation;

$$\begin{aligned} X_{ij} \sim \text{Poisson}(\lambda_{ij}) \quad & \& \quad Y_{ij} \sim \text{Poisson}(\mu_{ij}) \\ \text{With:} \quad \log(\lambda_{ij}) = \alpha_i - \beta_j + \eta \quad & \& \quad \log(\mu_{ij}) = \alpha_j - \beta_i \end{aligned} \tag{3}$$

Where, in Equation (3), $\alpha_k$ represents the AS of team $k$, with $k \in (i, j)$, $\beta_k$ similarly represents DS of team $k$, with $k \in (i, j)$, and $\eta$ is representative of the common HGA for each team.

This model (as well as Model 2 and Model 3) have a corresponding likelihood formula of;

$$\mathcal{L}(\boldsymbol{\lambda}, \boldsymbol{\mu}) = \prod_{i=1}^{n} \prod_{j=1}^{n} \frac{\lambda_{ij}^{x_{ij}} e^{-\lambda_{ij}}}{x_{ij}!} \times \frac{\mu_{ij}^{y_{ij}} e^{-\mu_{ij}}}{y_{ij}!}$$

With $\boldsymbol{\lambda} = (\lambda_{11}, \lambda_{12}, \ldots, \lambda_{nn})$, $\boldsymbol{\mu} = (\mu_{11}, \mu_{12}, \ldots, \mu_{nn})$, the parameter vectors for the model, and $x_{ij}$ and $y_{ij}$ the count data for number of home and away goals.

#### 3.2.2 Model 2

The second model we posit uses the same response variables as Model 1. In contrast to Model 1 however, Model 2 instead utilises Ability parameters in conjunction with individual HGA parameters as explanatory variables.
This model is formulated as follows;

$$\begin{aligned} X_{ij} \sim \text{Poisson}(\lambda_{ij}) \quad & \& \quad Y_{ij} \sim \text{Poisson}(\mu_{ji}) \\ \text{With:} \quad \log(\lambda_{ij}) = \sigma_i - \sigma_j + \eta_i \quad & \& \quad \log(\mu_{ij}) = \sigma_j - \sigma_i \end{aligned} \tag{4}$$

Where in Equation (4) $\sigma_k$ represents the Ability of team $k$, with $k \in (i, j)$, and $\eta_i$ represents the HGA of the $i^{th}$ team.

#### 3.2.3 Model 3

Our third model is an amalgamation of both models 1 and 2. Model 3 uses AS and DS parameters (as in Model 1), as well as individual HGA parameters (as in Model 2) as explanatory varianbles.
Thus, Model 3 is formulated as below;

$$\begin{aligned} X_{ij} \sim \text{Poisson}(\lambda_{ij}) \quad & \& \quad Y_{ij} \sim \text{Poisson}(\mu_{ji}) \\ \text{With:} \quad \log(\lambda_{ij}) = \alpha_i - \beta_j + \eta_i \quad & \& \quad \log(\mu_{ij}) = \alpha_j - \beta_i \end{aligned} \tag{5}$$

with the parameters defined similarly to the corresponding parameters in Model 1 and Model 2.

### 3.3 Validation

Proving the validity of assumptions 1 and 2 listed within Section 3.1 is done via diagnostics performed post-fitting of the model. Assumptions 4 and 5 are satisfied by the model formulation, with linear predictors $\alpha_i$ - $\beta_j + \eta$ and $\alpha_j$ - $\beta_i$ for Model 1, with analogous predictors for models 2, 3, 4 and 5.
The link function, $\boldsymbol{g}(\cdot)$, in each case is a logarithmic one ($\boldsymbol{g}(\mu) = \log(\mu) = \psi$), which by definition is both smooth and invertible, as required.

The proof of the satisfying of Assumption 3 can be done analytically by proving that a random variable following a Poisson distribution is part of the exponential family of distributions, as below.
If we take random variable Y $\sim$ Poisson($\mu$), this has the probability density function $f(y|\mu)$, where

$$f(y|\mu) = \frac{\mu^y e^{-\mu}}{y!}$$

which can be re-formulated to satisfy Equation 1 via the following transformations;

$$\begin{aligned} f(y|\mu) &= e^{y\log(\mu)} e^{-\mu} e^{-\log(y!)} \\ &= \exp\left[y\log(\mu) - \mu - \log(y!)\right] \end{aligned}$$

Letting $\theta = \log(\mu)$

$$\begin{aligned} f(y|\theta) &= \exp\left[y\theta - e^{\theta} - \log(y!)\right] \\ &= \exp\left[\frac{y\theta - e^{\theta}}{1} - \log(y!)\right] \end{aligned}$$

Which clearly satisfies Equation 1, with $\phi = 1$, k$(\theta) = e^{\theta}$ and c$(y,\phi) = \log(y!)$ as required.

## 3.4 Fitting

### 3.4.1 Initial Specifications

When fitting each of the three aforementioned univariate Poisson models within R version 1.1.442, we are presented with the issue of parameter identifiability.[2]
This issue is raised as the number of parameter values which maximise the likelihood for each model (AS, DS, ability and HGA) is effectively infinite, thus compromising the ability to find singular parameter estimates for the models. Appropriate constraints on the parameters help negate this problem.

There are a few possibilities for the choice of contraints to place upon the parameters in order to achieve parameter identifiability. [2] Cattelan et. al (2013) raises the option of placing a summation constraint upon the abilities of each team, such that the sum of all the abilities is equal to zero. For example, in Model 2, we would have: $\sum_{k=1}^{20} \sigma_k = 0$.
Instead, we choose a simpler constraint, selecting (again using Model 2 as an example) one team's ability, subsequently setting this ability equal to 0, and using this as a reference point to compare the remaining teams' parameter estimates to.
In this case we select Arsenal as the reference point, as they are the first team alphabetically and feature in every Premier League season within the dataset, aiding comparability.

### 3.4.2 Model 1

Making use of R's optim function in order to numerically maximise the likelihood for Model 1, we obtain parameter estimates for the log Attack Strength (AS), and Defensive Strength (DS), which we convert to the 'goals' scale by exponentiating. We also utilise the delta method in order to find standard errors for these parameters. We observe parameter estimates for the AS and DS of each of team within the dataset in the $8^{th}$ year of study (used as an example) represented graphically within Figure 3, below.

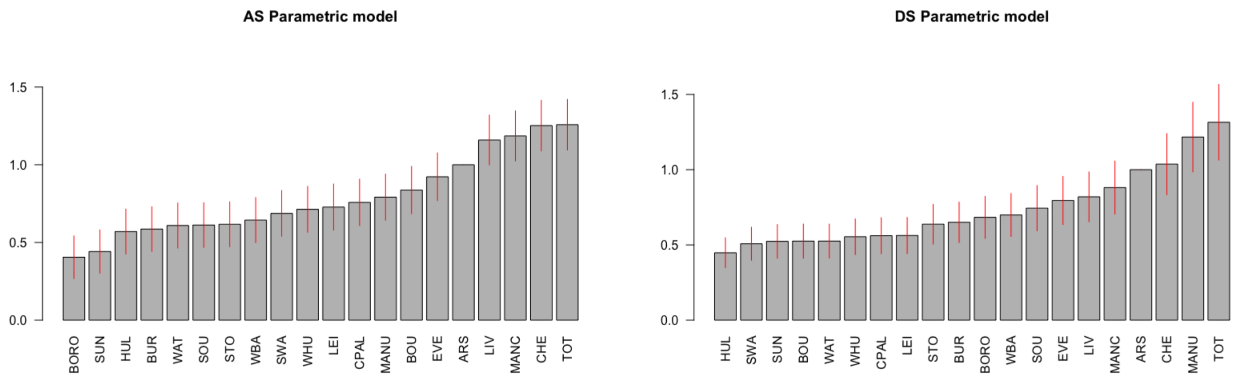



Figure 3: AS and DS estimates for each team in the $8^{th}$ season of study relative to Arsenal from Model 1. Red bars indicate standard errors of the parameter estimates.

From these estimates, it would appear that Tottenham had the highest AS *and* the highest DS within the $8^{th}$ season of study, with Hull City and Middlesbrough having the lowest DS and AS respectively. Upon viewing these results in comparison with the number of goals scored and conceded by each team within the corresponding season, this model appears to agree with the reality of the season. Tottenham scored the highest number of goals within the season (86) and conceded the least (26), whilst Middlesbrough scored the lowest number of goals (27). Similar trends appear to be present for the remainder of the teams in question. This indicates that Model 1 seems to be an adequate fit for the data.

Table 1: HGA scores common to each team, and their respective standard errors in each year of study

| **Season** | **HGA parameter Estimate** | **Standard Error** |
|---|---|---|
| 2009/10 | 1.602 | 0.0791 |
| 2010/11 | 1.403 | 0.0728 |
| 2011/12 | 1.340 | 0.0710 |
| 2012/13 | 1.267 | 0.0687 |
| 2013/14 | 1.321 | 0.0706 |
| 2014/15 | 1.362 | 0.0747 |
| 2015/16 | 1.239 | 0.0691 |
| 2016/17 | 1.356 | 0.0715 |
| 2017/18 | 1.374 | 0.0737 |

In addition to the AS and DS estimates, the common HGA parameter obtained for the $8^{th}$ year of study was found to be 1.356.
The remaining years' estimates are found within Table 1, and are fairly consistent with the score for year 8, indicating that HGA is a parameter that should be accounted for, as discussed in [3] Norman & Clarke.

### 3.4.3 Model 2

Fitting analogously the second model under consideration within our study (including Ability of each team, and a HGA which varies by team), again exponentiating to produce estimates for Ability and HGA on the 'goals' scale and producing standard errors via the delta method, we produce the estimates represented within Figure 4.

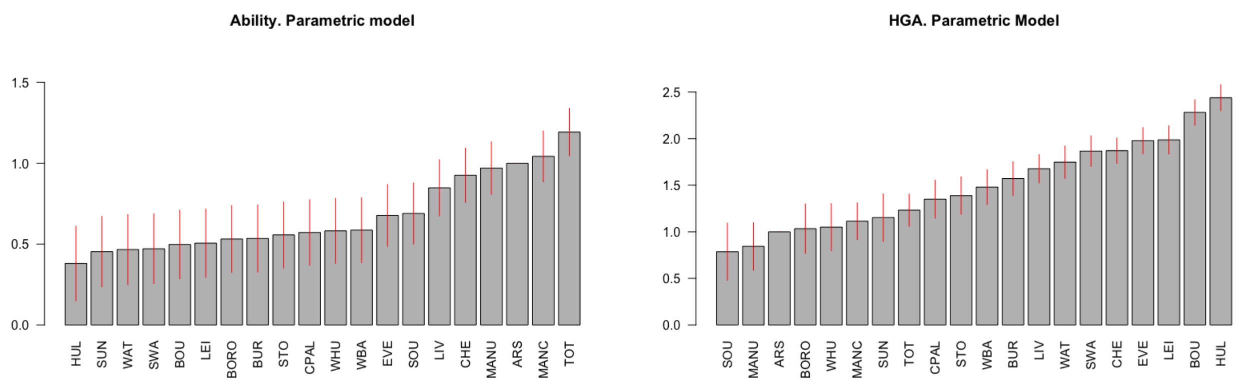



Figure 4: Ability and HGA estimates for each team in the $8^{th}$ season of study relative to Arsenal from Model 2. Red bars indicate standard errors of the parameter estimates.

We see from this Figure that the team with the highest Ability score within the dataset appears to be Tottenham. This is intuitive, as Ability accounts for both goals scored, and goals conceded, and as we saw from the results obtained for Model 1, Tottenham conceded the least goals and scored the most within the season under study. However, the results also may show that this particular model may be unsuitable for use, as actual eventual champions Chelsea are ranked $4^{th}$ in terms of Ability, and an accurate model would intuitively rank the team that scored the highest number of points throughout the season as being the highest in terms of Ablility.

With regards to HGA, the teams' parameter estimates correspond to the percentage of each team's points that were accrued at home throughout the season. Accurately here, Hull have the highest HGA of any team within the league, which is reasonable, as they gained their highest percentage of points at home within the season (82.4%). Conversely, Southampton have the lowest HGA parameter produced by Model 2. This does not seem to be representative of the data for the season, as Manchester United for example have a lower home win percentage (50%).

The nature of the parameter rankings for Model 2 therefore seem to indicate that this model appears to be an inadequate fit for the data.

### 3.4.4 Model 3

Finally, fitting our third model, including AS, DS and a individual HGAs, we produce parameter estimates for each team, seen within Figure 5 below.

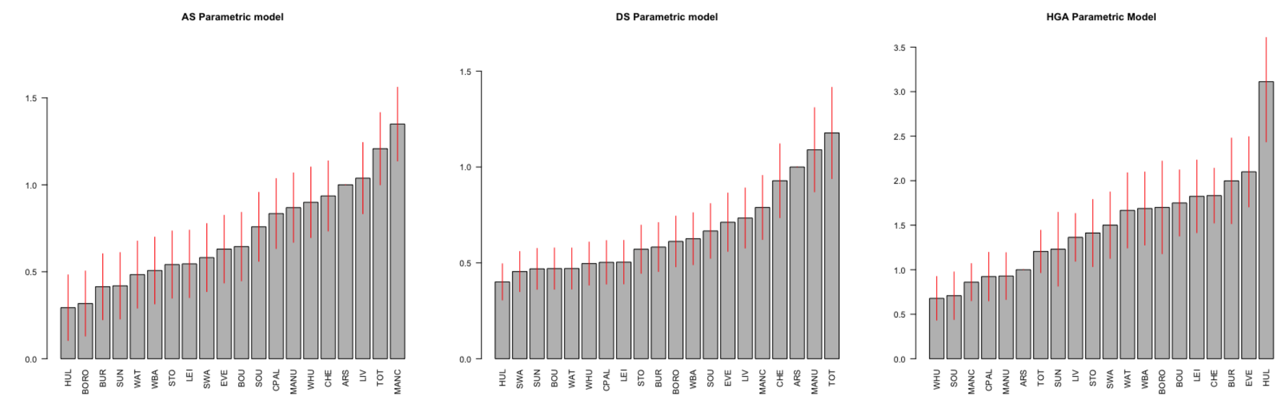



Figure 5: AS, DS and HGA estimates for each team in the $8^{th}$ season of study relative to Arsenal from Model 3. Red bars indicate standard errors of the parameter estimates.

The parameter estimates for the AS and DS within this model appear to differ significantly from those found to be accurate within Model 1. They no longer correspond to goals scored and conceded.
A similarly poor result is obtained with respect to the HGA parameters, with West Ham United this time obtaining the lowest score, despite winning 60.5% of their points at home as opposed to Manchester United for example (ranked $16^{th}$), who won 50% of their points at home. Thus, neither the HGA, AS or DS parameters appear to be accurate within the context of the season's results.

The poor parametric results produced by this model are indicative of poor fit, possibly due to overfitting within the model. This, along with the results for the first two models, will be evaluated mode formally via the observation of AIC and BIC scores for each model.

## 3.5 AIC

In order to acquire an initial numerical estimate of the relative quality of each univariate model utilised within our fitting, we make use of the Akaike information criterion (AIC) [1] value for each in turn.

$$\mathbf{AIC} = -2l(\boldsymbol{\theta}) + 2\mathrm{K} \tag{6}$$

The AIC score of a model is defined within Equation 6 above, in which K represents the number of parameters within each model, and $l(\boldsymbol{\theta})$ represents the log likelihood of the models' parameters (where $\boldsymbol{\theta}$ is representative of the vector of parameters, for example in Model 1, $\boldsymbol{\theta} = (\alpha_k, \beta_k, \eta)$, for k $\in$ ($i$,$j$).)

The AIC statistic seeks to find the model which is the best fit for the model, with the fewest number of parameters (i.e. the least complex). The model with the lowest AIC is the optimal relative model, with the value of K (the number of parameters, and hence the complexity) acting as a penalty for the complexity of the model, and the likelihood being involved as an indicator of goodness of fit.

We proceed to assessing each model via their relative AIC. The AIC for each model in each season of study can be found within Table 2.

Table 2: AIC values for all models in every season (lowest AIC model highlighed in blue).

| | **Model Number** | | |
|---|---|---|---|
| **Season** | 1 | 2 | 3 |
| 2009/10 | 2194.8 | 2203.1 | 2217.6 |
| 2010/11 | 2245.6 | 2250.5 | 2265.3 |
| 2011/12 | 2261.4 | 2277.9 | 2283.8 |
| 2012/13 | 2260.4 | 2285.3 | 2281.2 |
| 2013/14 | 2254.3 | 2277.5 | 2282.5 |
| 2014/15 | 2177.4 | 2184.9 | 2201.4 |
| 2015/16 | 2243.4 | 2246.6 | 2258.9 |
| 2016/17 | 2225.9 | 2222.0 | 2231.6 |
| 2017/18 | 2192.1 | 2220.2 | 2224.9 |

As we would expect, as Model 3 is by far the most complex in terms of number of parameters (Model 3 possesses 57 parameters, as opposed 39 for Model 1 and 40 for Model 2), Model 3 has the highest AIC value for each season. This is due to the penalisation from the complexity parameter, K.
Models 1 and 2 however are very similar in terms of complexity, and this is reflected in the results, with these two models always producing the lowest AIC values (Model 1: 9 times, Model 2: 1 time. This suggests that including AS and DS parameters in conjunction with a common HGA parameter produces a model with a better fit than if Ability parameters and individual HGA parameters are used (as is the case in Model 2). The AIC results appear to be conclusive as to which parametrisation to use within the model. This will be further explored via the analysis of BIC values.

## 3.6 BIC

BIC [12] is similar to AIC in its usage, but carries a different penalisation value. Where the formula for AIC (Equation 6) makes use of 2K, BIC uses Klog(n), so is formulated as in Equation 7.

$$\mathbf{BIC} = -2l(\boldsymbol{\theta}) + \mathrm{K}\log(n) \tag{7}$$

With respect to BIC, we observe from Table 3 that, similar to the result for the AIC (S3.5), Model 3 is likely penalised to severely by its complexity to be of a good fit to the data, resulting in a high BIC value for every year of study.
As was the case with the AIC results, Model 1 and Model 2 both have years in which they

Table 3: BIC values for all models in every season (lowest BIC model highlighed in blue).

| | Model Number | | |
|---|---|---|---|
| Season | 1 | 2 | 3 |
| 2009/10 | 2348.5 | 2352.8 | 2442.2 |
| 2010/11 | 2399.3 | 2400.2 | 2489.9 |
| 2011/12 | 2415.1 | 2427.6 | 2508.4 |
| 2012/13 | 2414.1 | 2435.1 | 2505.8 |
| 2013/14 | 2408.0 | 2427.3 | 2507.0 |
| 2014/15 | 2331.1 | 2334.7 | 2426.0 |
| 2015/16 | 2397.1 | 2396.4 | 2483.5 |
| 2016/17 | 2379.6 | 2371.8 | 2456.2 |
| 2017/18 | 2345.8 | 2369.9 | 2449.5 |

obtain the lowest BIC value, although with respect to BIC, Model 1 seems to be more consistently the best model to fit for the data, obtaining the lowest BIC value in 7 of the 9 years within the study. Model 2 is only the closest fit to the data in consecutive 2015/16 and 2016/17 seasons.

On the balance of both the AIC and BIC scores then, it would appear that of the univariate Poisson models under consideration, Model 1 (which included AS and DS scores as well as a common HGA parameter) appears to be the best fit for the years under study. The true utility of this model will be determined by diagnostics.

## 3.7 Diagnostics

Within the prior sections, we assessed the relative goodness of fit of each of our three univariate Poisson models via the comparison of their corresponding AIC and BIC values. These tests were conclusive in presenting Model 1 as the superior model of the three univariate Poisson models.
However, although these measures give insight as to which model is the best fit in comparison to the others, this measure is rather subjective, and does not give an indication of how truly adequate each is.

In order to obtain a more objective measure of Model 1's fit for the data in question, we turn to diagnostic analyses, largely comprised of inferences made on three forms of residuals. These diagnostic tests facilitate the assessment of whether the underlying assumptions of our chosen superior model (that our observations are independent and Poisson-distributed) are satisfied with respect to the data. The satisfaction of these assumptions by a model would indicate that said model is of adequate fit.
We also use diagnostics in order to check our data for outlying entries, which (if systematic) may be indicative of poor model fit.

Within Sections 3.7.1, 3.7.2 and 3.7.3, we look at the three forms of residuals used within our diagnostic processes for our model, and in Section 3.7.4, we assess the uses of these residuals in more detail.

### 3.7.1 Raw Residuals

In order to assess the underpinning assumptions of the model, we must make use of fitted values, which are the 'expected' response variable produced by each model.
These fitted values are used to produce residual values for each point in the model via the formula found within Equation 8.

$$e_{ij} = X_{ij} - \hat{\lambda_{ij}} \tag{8}$$

In Equation 8, $e_{ij}$ is the residual value corresponding to a match between home team $i$ and away team $j$, $X_{ij}$ the observed value of the response variable (number of home goals) and $\hat{\lambda_{ij}}$ is the fitted (expected) value of the response obtained from the corresponding model. We have a similar equation for $Y_{ij}$ and $\hat{\mu_{ij}}$ (away goals).

Assessing the distribution and properties associated with these residuals enables us to assess the satisfaction of the underpinning assumptions of our model, which in turn allows use to make inferences upon its objective utility in modelling the data at hand.

### 3.7.2 Pearson Residuals

In addition to raw residuals, we also use Pearson residuals within our diagnostic process. These are formulated (as described by [6] Dobson) within Equation 9.

$$r^{x_{ij}} = \frac{e_{ij}}{\sqrt{\hat{\lambda_{ij}}}}, \qquad r^{y_{ij}} = \frac{e_{ij}}{\sqrt{\hat{\mu_{ij}}}} \tag{9}$$

Where $r^k$ represents the residuals related to the home ($k = x_{ij}$) or away ($k = y_{ij}$) goals within a matchup between home team $i$ and away team $j$, and the remaining terms are defined as previously within this paper.

### 3.7.3 Deviance Residuals

Finally, as well as analysing the raw and Pearson forms of residuals within our diagnostic process, we also analyse deviance residuals. The formula for these residuals involves the log of the correspsonding probability distribution related to the model, evaluated at both the response variable measures and the fitted values produced by the model.
As an example, we take the Poisson model for the home goal count, shown within Equation 3. Within this equation, our response variables are denoted by $x_{ij}$, and our fitted values by $\hat{\lambda_{ij}}$. We then utilise Equation 10

$$d_{ij} = 2\phi \left\{ log\{f(x_{ij}|x_{ij},\phi)\} - log\{f(x_{ij}|\hat{\lambda_{ij}},\phi)\} \right\} \tag{10}$$

Where in Equation 11, $f(\cdot)$ represents the probability distribution of the model, and $\phi$ the model's Scale parameter.
For a model utilising a Poisson distribution (as is the case with our model), $\phi = 1$ (as shown in Section 3.3) and the formula for $d_i$ is as displayed in Equation 11.

$$d_{ij} = 2 \left\{ x_{ij} log \left\{ \frac{x_{ij}}{\hat{\lambda_{ij}}} \right\} - e_{ij} \right\} \tag{11}$$

Where in Equation 11, $e_{ij}$ is as in Equation 8. The $i^{th}$ deviance residual is then found via the formula: $sign(x_{ij} - \hat{\lambda_{ij}})\sqrt{|d_{ij}|}$.

#### 3.7.4 Residual Uses

As mentioned previously, the residuals of the forms described within Sections 3.7.1, 3.7.2 and 3.7.3 may be used to check model assumptions and the properties of the model at hand.

The first assumption we test for our model is the assumption that the observations made are independently distributed of one another. This assumption is tested using the Pearson residuals for the model, described wihtin Section 3.7.2.
The squared Pearson residuals for the model follow a Chi-squared ($\chi^2_{n-p}$) distribution, which can then be tested for independence via analysis of Chi-squared $p$-values.

Another diagnostic role of the residuals is assessing for outliers. An abundance of outlying data entries which do not fall in line with the values expected from the model may be indicative of an ill-fitting model.
This assessment done is by the analysis of the raw residuals via the utilisation of Cook's distance (described within Section 3.7.6).

One final assumption we may assess the validity of in order to gauge the goodness of fit for our model is the assumption of variance homogeneity across assessment groups. This allows the assessment of whether or not the data follows the assumed distribution (Poisson). These tests for variance homogeneity are done by analysing the variance of the deviance residuals (from Section 3.7.3) for our model.
Analysing the variance of these residuals for different groups within the model is indicatory of the variance of our parameter estimates. Therefore, if these variances are fairly homogeneous, we have a fairly good indication that the variance homogeneity assumption for our model is satisfied. Thus, that the distribution of the data is seen to be suffiently close to the Poisson model assumed by our model.

If the validity of all these assumptions is demonstrated via the diagnostic processes listed above, then we may conclude that our model is of an adequate fit for the data.

### 3.7.5 Independence

Table 4: Squared sum of Pearson's residuals and Chi-squared $p$-values for models 1-4.

| Model | Parameter | Sum of Pearson's Residuals | Chi-Squared $p$-value |
|---|---|---|---|
| 1 | $\lambda_{ij}$ | 340.234 | 0.502 |
| | $\mu_{ij}$ | 363.555 | 0.192 |
| 2 | $\lambda_{ij}$ | 378.398 | 0.074 |
| | $\mu_{ij}$ | 387.977 | 0.037 |
| 3 | $\lambda_{ij}$ | 468.517 | 2.053 x $10^{-7}$ |
| | $\mu_{ij}$ | 340.282 | 0.244 |
| 4 | $\lambda_{ij}$ | 356.469 | 0.259 |
| | $\mu_{ij}$ | 386.426 | 0.045 |

Calculating the Pearson's residuals and the Chi-squared test statistics for each model (as described within Section 3.7.4), we produce the results found within Table 4.
We see from Table 4 that only Model 1 produces a $p$-value of $>$ 0.05 for both of $\lambda_{ij}$ and $\mu_{ij}$, thus passing the Chi-squared test for independence of residuals.
This finding, paired with the results of the AIC and BIC tests, indicate that a model (such is the case in Model 1) including AS and DS parameters, as well as a common HGA parameter passes the test for independent observations, and thus appears to be a good fit for the data when judging on this criteria.

### 3.7.6 Outliers

In order to identify outlying points within the dataset with respect to the model, (as noted within Section 3.7.4) we turn to the analysis of [4] Cook's Distance, the formula for which can be found within Equation 12.

$$D_{ij} = \frac{e_{ij}^2}{s^2 p}\left[\frac{h_{ij}}{(1-h_i)^2}\right] \tag{12}$$

Within Equation 12, $e_{ij}$ is defined as before, $s^2$ represents the mean squared error, $p$ the number of parameters within the model (in Model 1, $p = 39$), and $h_{ij}$ the $(i, j)^{th}$ entry of the model's hat matrix. $D_{ij}$ values of greater than $8/(n\text{-}2p)$, where $n$ is the number of data points, are considered to be outlying, or 'of high influence' (Cook, 1977) [4].

The $D_{ij}$ values for each data point within the $8^{th}$ year of study produced by Model 1 are plotted within Figure 6.
From this plot, we see that of the 760 residual entries (380 games, AS and DS parameters within each game), 6 may be considered to be outlying (above the $8/(n\text{-}2p)$ threshold).

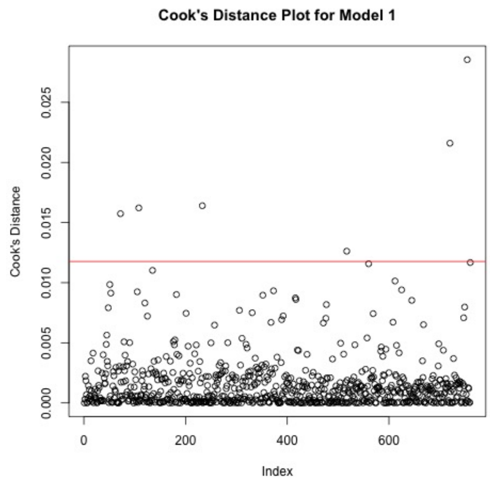



Figure 6: Cook's Distance plot of the deviance residuals for Model 1. Red line is plotted at $8/(n\text{-}2p)$

Corresponding to the points in question, we have the games; Bournemouth 6-1; Hull Liverpool; 6-1 Watford Everton 6-3 Bournemouth; West Ham 1-5 Arsenal; Leicester 1-5 Tottenham and Hull 1-7 Tottenham.
All of the points of high influence thus correspond to very high-scoring games (games which feature 6 or more goals), which may indicate that our superior model (Model 1) underestimates the propensity for Premier League teams to be involved in high-scoring games.

However, despite each outlying point being related to high-scoring games, not every game of this nature was found to be an outlier. There were 22 of such games, which means only 27% of the very high scoring games were associated with outlying points.
Also, in the context of 380 games played within the season, these 6 points of high influence seem insignificant, and still point to Model 1 being adequate for the data at hand due to the rarity of the very high-scoring games.

### 3.7.7 Overdispersion

Another assumption which we test the validity of is that there exists no evidence of overdispersion of the data relative to the Poisson model.
This problem occurs when the data in question has a greater variance value than that accounted for within the model. This issue is particularly prevalent within Poisson formulations, as the variance of the model is determined by its mean, resulting in inflexibility.

We need not analyse residuals in order to assess the existence of overdispersion within this dataset. In order to diagnose possible overdispersion within our univariate Poisson model (Model 1), we instead may apply a Negative Binomial model to the data, which utilises an extra dispersion parameter to account for any additional variation within the dataset.
In the case of overdispersion, fitting a Negative Binomial model to the dataset in question (although not changing the parameter estimates) would result in an increase in standard

Table 5: AS and DS Standard error scores for a selection of teams for both Poisson (PSE) and Negative Binomial (NBSE) Models in the $8^{th}$ season of study

| **Team** | **Parameter** | | | |
|---|---|---|---|---|
| | **AS** | | **DS** | |
| | **PSE** | **NBSE** | **PSE** | **NBSE** |
| BOU | 0.176 | 0.185 | 0.180 | 0.189 |
| CHE | 0.171 | 0.180 | 0.192 | 0.200 |
| EVE | 0.181 | 0.190 | 0.190 | 0.198 |
| LEI | 0.183 | 0.191 | 0.183 | 0.192 |
| MANC | 0.184 | 0.193 | 0.175 | 0.184 |
| BORO | 0.201 | 0.209 | 0.201 | 0.209 |
| STO | 0.191 | 0.199 | 0.189 | 0.197 |
| SWA | 0.178 | 0.187 | 0.184 | 0.192 |
| WAT | 0.184 | 0.193 | 0.184 | 0.193 |
| WHU | 0.188 | 0.196 | 0.178 | 0.187 |

error for the parameters.

Within Table 5, we observe the standard error scores with respects to both the Poisson and Negative Binomial models for a selection of teams within the $8^{th}$ season of study.
These results illustrate that every team's parameter estimates have lower standard error values when using the Poisson model than when using the Negative Binomial model. Thus, we conclude that there exists some evidence of overdispersion in the model.
The increases in standard error however, are minimal. To investigate how the minor overdispersion found affects the model fit, we turn to observing the AIC scores for both Negative Binomial and Poisson formulations.

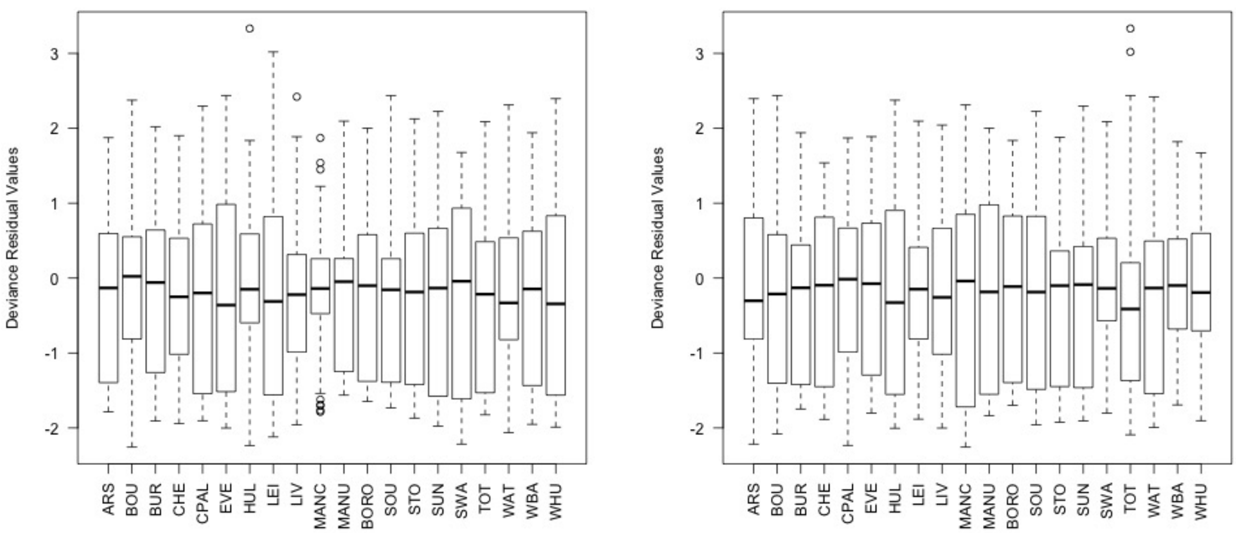
Table 6: AIC scores for Poisson and Negative Binomial formulations of Model 1 for each season of study. Lowest score highlighted in blue.

| **Season** | **Poisson AIC** | **NegBin AIC** |
|---|---|---|
| 2009/10 | 2194.8 | 2196.8 |
| 2010/11 | 2245.6 | 2247.6 |
| 2011/12 | 2261.4 | 2263.4 |
| 2012/13 | 2260.4 | 2262.4 |
| 2013/14 | 2254.3 | 2256.3 |
| 2014/15 | 2177.4 | 2179.4 |
| 2015/16 | 2243.4 | 2245.4 |
| 2016/17 | 2225.9 | 2227.9 |
| 2017/18 | 2192.1 | 2194.1 |

Upon fitting the two models (Model 1 in both Poisson and Negative Binomial formulation) we obtain the AIC values listed within Table 6. From this we see that the fitting of the Negative Binomial model does not improve model fit with respect to the data in any of the nine seasons. Thus, we do not deem the overfitting present within the dataset to be significant.

### 3.7.8 Homogeneity of Variance

The 'groups' that we make use of to test the assumption of variance homogeneity may simply be the different teams under study within each respective season, and in order to get an understanding as to whether the variances of these 'groups' are homogeneous, we may first make use of box plots pertaining to the deviance residuals (seen within Figure 7).




(a) AS (b) DS

Figure 7: Bar charts for the deviance residuals of each teams (a) AS and (b) DS parameters, grouped by attacking and defending team

From the boxplots, it can be drawn that the residual variances seem relatively uniform, though there are a few possible minor deviations from this. For example, in the case of the attacking strength for Manchester City '(MANC)'. The variance for this parameter's residuals is comparatively small when viewed alongside the other AS parameter estimates, as well as it possessing a few outlying residual values.

There also appear to be some outliers for the AS parameters of Hull ('HUL') and Liverpool ('LIV') as well as the DS parameter for Tottenham ('TOT'), though the variances for these appear to be in line with those of the other teams in question.

As a formal assessment of the variances, we conducted Levene's tests upon the model, again using the different teams' sets of parameter estimates as the groups under assessment for variance homogeneity.

For the AS parameters of each team under study, the respective Levene's test produced a test statistic of 1.064, with a corresponding $p$-value of 0.384. The Levene's test with respect to the DS parameters produced a test statistic of 0.957, with a corresponding $p$-value of 0.512.

Thus, from these formal assessments of variance homogeneity, as each $p$-value is $>0.05$, at a 5% level we may conclude that there is no evidence of variance heterogeneity amongst the parameter estimates, as assumed. Again, these tests point to the Univariate Poisson model utilising AS, DS and a common HGA parameter being a good fit for the data at hand.

## 3.8 Correlation of Home and Away Goals

As well as the model assumptions assessed within the previous subsections, within the previously proposed models, it was assumed that the home and away goals in each game follow Poisson distributions which are *independent* of each other.
However, this assumption may prove to be invalid, possibly due to the nature of the sport being played. In football, teams interact with one-another which would lead one to believe that the scoring of one or more goals by one team has a significant effect on the possibility of the opposing team scoring; i.e. the two goal counts are *correlated*.

Possible correlation between these two sets of variables has previously been displayed within the paper by [8] Lee (1997), which was mentioned in the subsequent paper by [7] Karlis & Ntzoufras (2003). Karlis & Ntzoufras stated in particular that correlation may come into play within the sport of basketball, due to the 'sequential' nature of the scoring attempts made by each team, with 'the speed of the game of one team leads to more opportunities for both teams to score.' The case for correlation between outcome variables assumes that the correlation is logical due to the interaction between teams.

We investigate the validity of this assumption via the analysis of correlation values for the previously-defined Pearson's residuals.
In Table 7, we observe the correlation between the Pearson's residuals for the $\lambda_{ij}$ and $\mu_{ij}$ variables for each model. These correlation values are calculated using every game within every year of study 2009/10-2017/18.

Table 7: Pearson's residual correlation and corresponding $p$-values for each year for each model.

| | Model Number | | | | | |
|---|---|---|---|---|---|---|
| | 1 | | 2 | | 3 | |
| Season | Corr($r^{x_{ij}}$, $r^{y_{ij}}$) | $p$-value | Corr($r^{x_{ij}}$, $r^{y_{ij}}$) | $p$-value | Corr($r^{x_{ij}}$, $r^{y_{ij}}$) | $p$-value |
| 2009/10 | -0.031 | 0.550 | -0.042 | 0.410 | -0.031 | 0.545 |
| 2010/11 | 0.039 | 0.446 | -0.013 | 0.805 | 0.003 | 0.957 |
| 2011/12 | -0.019 | 0.714 | -0.031 | 0.550 | -0.036 | 0.488 |
| 2012/13 | -0.006 | 0.900 | -0.004 | 0.937 | -0.028 | 0.582 |
| 2013/14 | 0.024 | 0.636 | -0.007 | 0.893 | -0.017 | 0.746 |
| 2014/15 | -0.005 | 0.921 | -0.047 | 0.362 | -0.043 | 0.401 |
| 2015/16 | -0.021 | 0.680 | -0.035 | 0.498 | -0.028 | 0.580 |
| 2016/17 | 0.012 | 0.822 | -0.005 | 0.917 | 0.015 | 0.772 |
| 2017/18 | -0.025 | 0.625 | 0.064 | 0.215 | 0.005 | 0.919 |

From Table 7, we see a mix of positive and negative correlation exhibited between the home

and away goal values.
However, using Pearson's product moment correlation coefficient to perform t-tests on the correlation values for each model, we see insufficient evidence of correlation between home and away goals in any model and in any year, leading us to believe the inclusion of a correlation parameter would not prove helpful within model fitting for the full data.

Although no correlation was found when the entire dataset was tested, [5] Dixon & Coles, as well as [7] Karlis & Ntzoufras highlight low scoring games as being particularly highly correlated with respect to number of home and away goals.
With this theory in mind, we assess the correlation for low-scoring games (0-0, 1-1, 1-0, 0-1) analogously for the same seasons, the results of this are found within Table 8.

Table 8: Pearson's residual correlation and corresponding $p$-values for each year for each model in low-scoring games. (significant $p$-values highlighted in blue.)

| | Model Number | | | | | |
|---|---|---|---|---|---|---|
| | 1 | | 2 | | 3 | |
| Season | Corr($r_i$, $r_i$) | $p$-value | Corr($r_i$, $r_j$) | $p$-value | Corr($r_i$, $r_j$) | $p$-value |
| 2009/10 | -0.114 | 0.027 | -0.162 | 0.002 | -0.095 | 0.065 |
| 2010/11 | -0.051 | 0.320 | -0.099 | 0.053 | -0.054 | 0.293 |
| 2011/12 | -0.111 | 0.031 | -0.122 | 0.018 | -0.086 | 0.093 |
| 2012/13 | -0.144 | 0.005 | -0.167 | 0.001 | -0.116 | 0.023 |
| 2013/14 | -0.123 | 0.016 | -0.141 | 0.006 | -0.136 | 0.008 |
| 2014/15 | -0.047 | 0.356 | -0.091 | 0.076 | -0.065 | 0.207 |
| 2015/16 | -0.103 | 0.044 | -0.143 | 0.005 | -0.115 | 0.025 |
| 2016/17 | -0.077 | 0.134 | -0.080 | 0.121 | -0.056 | 0.278 |
| 2017/18 | 0.017 | 0.739 | 0.042 | 0.419 | 0.006 | 0.911 |

In contrast to the results found within the full dataset for each year, when examining the seasonal data for only the low-scoring games in the years under study, significant correlation is exhibited by the first 3 models in various seasons, as shown within Table 7. These correlation values are always positive, indicating that when a goal is scored the home team, the away team are more likely to follow suit.
The findings in terms of low-scoring games are consistent with those posited by [7] Karlis & Ntzoufras and [5] Dixon & Coles.

# 4 Bivariate Poisson Models

## 4.1 Karlis & Ntzoufras Model

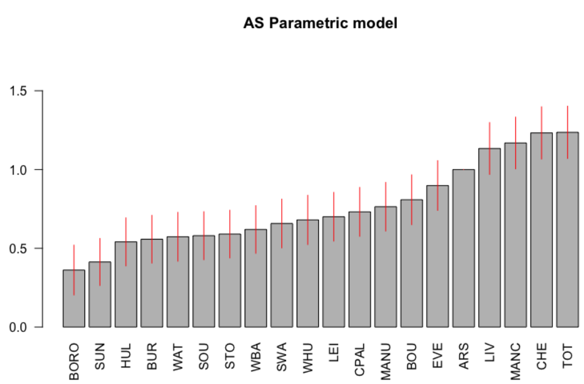
In an attempt to account for the effect of the correlation exhibited between home and away goal counts within the seasonal data, we must posit models which account for said correlation via the inclusion of additional parameters.

The first of such models is similar to the initial model proposed by [7] Karlis & Ntzoufras (2003). This model performs in a similar way to Model 1 and Model 3; including both AS and DS parameters. The model also utilises a common HGA parameter as is the case within Model 1.
In contrast to the previous models however, this model is a *bivariate* Poisson model, including an extra term common to both $\lambda_{ij}$ and $\mu_{ij}$, which we denote $\rho$.
This $\rho$ value is used to assess the correlation between the number of home goals within each game, $\lambda_{ij}$, and the number of away goals, $\mu_{ij}$. However, this model does not account for the possible nature of the correlation (e.g. correlation only occurring in low-scoring games, as suggested by Section 3.8). Hence, our new model (which we term 'Model 4') has a corresponding likelihood formula of;

$$\mathcal{L}(\boldsymbol{\lambda}, \boldsymbol{\mu}, \rho) = \prod_{i=1}^{n}\prod_{j=1}^{n}\left\{\frac{\lambda_{ij}^{x_{ij}}e^{-\lambda_{ij}}}{x_{ij}!} \times \frac{\mu_{ij}^{y_{ij}}e^{-\mu_{ij}}}{y_{ij}!} \times \sum_{k=0}^{min(x_{ij},y_{ij})} \binom{x_{ij}}{k}\binom{y_{ij}}{k}k!\left(\frac{\rho}{\lambda_{ij}\mu_{ij}}\right)^{k}\right\} \tag{13}$$

Where $\boldsymbol{\lambda} = (\lambda_{11},\lambda_{12},...,\lambda_{nn})$, $\boldsymbol{\mu} = (\mu_{11}, \mu_{12}, \ldots, \mu_{nn})$ are again the parameter vectors for the model, $(x_{ij}, y_{ij})$ are values of home and away goals within the data, and $k$ represents the minimum value of the home and away goals in each game.

#### 4.1.1 Parameter Estimates



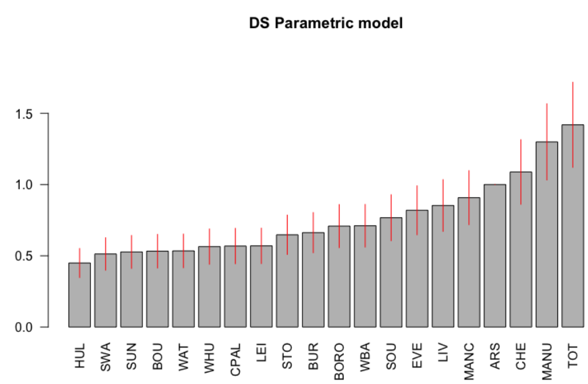




Figure 8: AS and DS estimates for each team in the $8^{th}$ season of study relative to Arsenal from Model 4. Red bars indicate standard errors of the parameter estimates.

Upon fitting the Karlis & Ntzoufras model (Model 4) to our data, we obtain the parameter estimates exhibited within Figure 8.
As expected due to their near-identical formulations, Model 4 produces very similar parameter estimates to that of Model 1. In particular, the ranking of each team's estimates remains the same between the two models. As a result of this, analogous conclusions to those drawn within Section 3.4.2. with regards to these estimates.

### 4.1.2 Assessment of Fit

As before, after examining the ranking of the parameter estimates, we move to examining the goodness of fit of Model 4 in comparison to those of the preceding models.

Observing the measures of both AIC and BIC for Model 4 similarly to models 1-3, we obtain the values displayed within tables 9 and 10.

Table 9: AIC values for all 4 models (univariate 1-3, bivariate 4) in every season (lowest AIC model highlighed in blue).

| | **Model Number** | | | |
|---|---|---|---|---|
| **Season** | 1 | 2 | 3 | 4 |
| 2009/10 | 2194.8 | 2203.1 | 2217.6 | 2192.4 |
| 2010/11 | 2245.6 | 2250.5 | 2265.3 | 2247.0 |
| 2011/12 | 2261.4 | 2277.9 | 2283.8 | 2263.2 |
| 2012/13 | 2260.4 | 2285.3 | 2281.2 | 2250.2 |
| 2013/14 | 2254.3 | 2277.5 | 2282.5 | 2255.7 |
| 2014/15 | 2177.4 | 2184.9 | 2201.4 | 2177.2 |
| 2015/16 | 2243.4 | 2246.6 | 2258.9 | 2242.1 |
| 2016/17 | 2225.9 | 2222.0 | 2231.6 | 2226.8 |
| 2017/18 | 2192.1 | 2220.2 | 2224.9 | 2193.4 |

With respect to AIC, in comparison to models 1-3 we see that Model 4 scores lowest in 4 of the 9 years of study. When considering BIC however, Model 4 only scores lowest in one of the 9 years.
However, as we see from the tables, Model 1 is seen to be the favourable model in 4 of the 9 years also with respect to AIC, and 6 of the 9 years when observing BIC scores.
When viewing the AIC and BIC results in conjunction with the scores for the other 3 models then, it would seem that including a correlation parameter between home and away goal counts which is not specific to the amount of goals scored within the game does not improve model fit. This is particularly evident when considering BIC scores.

Table 10: BIC values for all 4 models (univariate 1-3, bivariate 4) in every season (lowest BIC model highlighed in blue).

| | **Model Number** | | | |
|---|---|---|---|---|
| **Season** | 1 | 2 | 3 | 4 |
| 2009/10 | 2348.5 | 2352.8 | 2442.2 | 2350.0 |
| 2010/11 | 2399.3 | 2400.2 | 2489.9 | 2404.6 |
| 2011/12 | 2415.1 | 2427.6 | 2508.4 | 2420.8 |
| 2012/13 | 2414.1 | 2435.1 | 2505.8 | 2407.8 |
| 2013/14 | 2408.0 | 2427.3 | 2507.0 | 2413.3 |
| 2014/15 | 2331.1 | 2334.7 | 2426.0 | 2334.8 |
| 2015/16 | 2397.1 | 2396.4 | 2483.5 | 2399.8 |
| 2016/17 | 2379.6 | 2371.8 | 2456.2 | 2384.4 |
| 2017/18 | 2345.8 | 2369.9 | 2449.5 | 2351.0 |

These conclusions make sense in the context of the results found within Table 7, which illustrated that there was no evidence of correlation between home and away goals when the score was not specified.
This leads us to consider the formulation of a model in which our correlation parameter is specific to low-scoring games. This is the case in the model formulated by [5] Dixon & Coles.

## 4.2 Dixon & Coles Model

### 4.2.1 Formulation

The Karlis & Ntzoufras model, although accounting for possible correlation between home and away goal counts within each match, does not consider the nature of the occurrence conditions of this correlation.
Within Section 3.8, we demonstrated the existence (as postulated by Dixon & Coles) of correlation between home and away goal counts when considering only low-scoring games as opposed to every game within a season.

In an attempt to account for the correlation they postulated, which we have demonstrated within Section 3.8, we turn to the model formulated by Dixon & Coles, which we will term 'Model 5'.
This model is similar to that of Model 1, though we include a correlation function within the formulation of the likelihood, a representation of which is illustrated by the probability function put forward by Dixon & Coles as in Equation 14.

$$\mathcal{P}(X_{ij} = x, Y_{ij} = y) = \tau_{\lambda_{ij},\mu_{ij}(x,y)} \times \frac{\lambda_{ij}^{x_{ij}} exp(-\lambda_{ij})}{x_{ij}!} \times \frac{\mu_{ij}^{y_{ij}} exp(-\mu_{ij})}{y_{ij}!} \tag{14}$$

In which $\lambda_{ij}$, $\mu_{ij}$, $X_{ij}$ and $Y_{ij}$ are as in previously mentioned models, and $\tau_{\lambda_{ij}}, \mu_{ij}(x, y)$ is the correlation function for the model, defined as below.

$$\tau_{\lambda_{ij},\mu_{ij}}(x_{ij}, y_{ij}) = \begin{cases} 1 - \lambda_{ij}\mu_{ij}\rho & \text{if } x_{ij} = y_{ij} = 0 \\ 1 + \lambda_{ij}\rho & \text{if } x_{ij} = 0,\ y_{ij} = 1 \\ 1 + \mu_{ij}\rho & \text{if } x_{ij} = 1,\ y_{ij} = 0 \\ 1 - \rho & \text{if } x_{ij} = y_{ij} = 1 \\ 1 & \text{otherwise} \end{cases} \tag{15}$$

This results in a likelihood formula of;

$$\mathcal{L}(\boldsymbol{\lambda}, \boldsymbol{\mu}, \rho) \propto \prod_{i,j(i \neq j)}^{N} \tau_{\lambda_{ij},\mu_{ij}}(x_{ij}, y_{ij}) exp(-\lambda_{ij}) \lambda_{ij}^{x_{ij}} exp(-\mu_{ij}) \mu_{ij}^{y_{ij}} \tag{16}$$

As shown in Dixon & Coles (1997). Within Equation (16), as before $\boldsymbol{\lambda} = (\lambda_{11}, \lambda_{12}, \ldots, \lambda_{nn})$, and $\boldsymbol{\mu} = (\mu_{11}, \mu_{12}, \ldots, \mu_{nn})$ and the function $\tau_{\lambda_{ij},\mu_{ij}}(x_{ij}, y_{ij})$ is defined as in Equation (15).

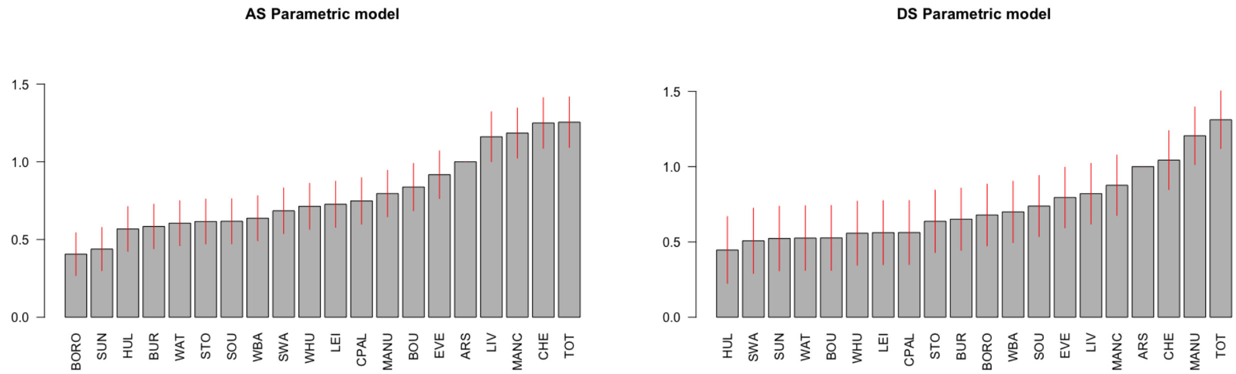



Figure 9: AS and DS estimates for each team in the $8^{th}$ season of study relative to Arsenal from Model 5. Red bars indicate standard errors of the parameter estimates.

#### 4.2.2 Parameter Estimates

When fitting the Dixon & Coles model (Model 5), we obtain the parameter estimates displayed within Figure 9 for each team within the season.
As was the case with Model 4, the formulation of Model 5 is near-identical to that of Model 1, except with the inclusion of a correlation function. This results in very similar parameter rankings to those produced by Model 1. However, conversely to the results within Model 4, there are slight differences within the rankings.
Within the rankings of the AS parameters, Stoke and Southampton swap positions from Model 1 and Model 5, with Stoke ranked $13^{th}$ and Southampton $14^{th}$ in Model 1, with the positions reversed in Model 5. However, this is of little consequence when viewed in the context of the 2016/17 season. In the season Southampton and Stoke scored an equal amount of goals (41), so it is logical that their AS parameters are interchangeable.
With respect to DS, the pairs of Watford and Bournemouth, and Leicester and Crystal Palace swap positions. Watford are ranked $15^{th}$ by Model 1, with Bournemouth $16^{th}$. Crystal Palace are ranked $13^{th}$ by Model 1, with Leicester $12^{th}$. The change of position between Crystal Palace and Leicester again is of little consequence, as both teams conceded 63 goals within the 2016/17 season. However, Bournemouth conceded 67 goals within the season, with Watford conceding 68, thus pointing towards Model 5's parameter rankings being slightly more accurate.

These parameter ranking results point marginally in favour of the Dixon & Coles model being the more accurate in the context of the 2016/17 season, though this will be investigated further by the observation of AIC and BIC scores for each model.

#### 4.2.3 Assessment of Fit

To measure adequacy of model fit within the Dixon & Coles model, as defined in the preceding part of Section 4.2, we again employ the measures of AIC and BIC and observe these in conjunction with the obtained results for the previously-studied models.

Table 11: AIC values for all 5 models (univariate 1-3, bivariate 4 and 5) in every season (lowest AIC model highlighed in blue).

| | **Model Number** | | | | |
|---|---|---|---|---|---|
| **Season** | 1 | 2 | 3 | 4 | 5 |
| 2009/10 | 2194.8 | 2203.1 | 2217.6 | 2192.4 | 2193.6 |
| 2010/11 | 2245.6 | 2250.5 | 2265.3 | 2247.0 | 2242.4 |
| 2011/12 | 2261.4 | 2277.9 | 2283.8 | 2263.2 | 2258.2 |
| 2012/13 | 2260.4 | 2285.3 | 2281.2 | 2250.2 | 2257.9 |
| 2013/14 | 2254.3 | 2277.5 | 2282.5 | 2255.7 | 2250.59 |
| 2014/15 | 2177.4 | 2184.9 | 2201.4 | 2177.2 | 2176.7 |
| 2015/16 | 2243.4 | 2246.6 | 2258.9 | 2242.1 | 2242.6 |
| 2016/17 | 2225.9 | 2222.0 | 2231.6 | 2226.8 | 2225.3 |
| 2017/18 | 2192.1 | 2220.2 | 2224.9 | 2193.4 | 2189.2 |

Table 12: BIC values for all 5 models (univariate 1-3, bivariate 4 and 5) in every season (lowest BIC model highlighed in blue).

| | **Model Number** | | | | |
|---|---|---|---|---|---|
| **Season** | 1 | 2 | 3 | 4 | 5 |
| 2009/10 | 2348.5 | 2352.8 | 2442.2 | 2350.0 | 2347.3 |
| 2010/11 | 2399.3 | 2400.2 | 2489.9 | 2404.6 | 2396.0 |
| 2011/12 | 2415.1 | 2427.6 | 2508.4 | 2420.8 | 2411.9 |
| 2012/13 | 2414.1 | 2435.1 | 2505.8 | 2407.8 | 2411.6 |
| 2013/14 | 2408.0 | 2427.3 | 2507.0 | 2413.3 | 2404.3 |
| 2014/15 | 2331.1 | 2334.7 | 2426.0 | 2334.8 | 2330.4 |
| 2015/16 | 2397.1 | 2396.4 | 2483.5 | 2399.8 | 2396.3 |
| 2016/17 | 2379.6 | 2371.8 | 2456.2 | 2384.4 | 2379.0 |
| 2017/18 | 2345.8 | 2369.9 | 2449.5 | 2351.0 | 2342.9 |

When observing the AIC and BIC values (displayed within Table 11 and Table 12), we see favourable scores for the Dixon & Coles model.
The model scores lowest with respect to AIC in 5 of the 9 years when compared to the previous 3 univariate models and the bivariate Karlis & Ntzoufras model. In addition to this, Model 5 scores lowest for BIC in 7 of the 9 years of study.
These two sets of scores for the model indicate strongly that the model formulated by Dixon & Coles produces a superior fit for the data in comparison to models 1-3 and the model put forward by Karlis & Ntzoufras (Model 4). Thus, it would appear that including a score-specific correlation function (which accounts for correlation in low-scoring games) marginally improves model fit.

# 5 Bayesian Approach Using Multiplicative Model

As we have displayed within the preceding sections, frequentist models reliant on maximum likelihood estimation perform sufficiently well in predicting the outcome of Premier League seasons.
However, what hasn't been investigated is how Bayesian methods may be employed to undertake the same task, and how these models perform.

## 5.1 Defining Terms

As in previous sections, within this section we define $i$ and $j$ to be the indices for the home and away team respectively in each match within a season. These are defined over $i, j \in \{1, \ldots, n\}$, with $n$ being the number of teams in the league under study (20 without constraints).
Similar to before also, we define $\alpha_i$ to be the AS parameter for the $i^{th}$ team, and $\beta_i$ to be the corresponding DS of the same team. We also define $\eta$ to be the HGA, as before.

Our model is formulated as follows;

$$\begin{aligned} X_{ij} \sim \text{Poisson}(\lambda_{ij}) \quad &\& \quad Y_{ij} \sim \text{Poisson}(\mu_{ij}) \\ \text{With:} \quad \lambda_{ij} = \alpha_i \beta_j \eta \quad &\& \quad \mu_{ij} = \alpha_j \beta_i \end{aligned} \tag{17}$$

Where again, $X_{ij}$ and $Y_{ij}$ represent the number of home and away goals respectively.

This model is similar to the our previously defined frequentist model, Model 1, from the preceding sections, which was found to be the optimal univariate Poisson model. We use this template for simplicity, as imposing a correlation function for the parameters in a Bayesian setting may prove difficult.

## 5.2 Finding Expressions for Conditional Distributions

In order to perform the Gibbs sampler using Bayesian methods, we must first find expressions for the conditional distribution functions of each of our terms. In order to do this, we must first impose prior distributions upon each parameter, thus allowing us to find joint distributions. These priors are as follows;

$$\begin{aligned} \alpha_i &\sim \text{Gamma}(\delta, \delta) \\ \beta_i &\sim \text{Gamma}(\delta, \delta) \\ \eta &\sim \text{Gamma}(\delta, \delta) \end{aligned} \tag{18}$$

Much like in the frequentist models illustrated in previous sections, the issue of parameter identifiability is something we must contend with within the Bayesian setting. To negate the problem this time, we impose the constraints on the common prior distributions (15). Fixing a small numerical value upon the parameter $\delta$ and treating this as known within each of these priors removes the possibility of infinite parameter solutions to the posterior distributions, thus enabling parameter identification in a similar way to in the frequentist models.

The common prior distributions lead to the joint posterior formulae within (16) and (17) below;

$$\begin{aligned}\boldsymbol{\pi}(\alpha_i, \beta_j, \eta|\mathbf{x}) \propto \prod_{i\neq j}\{(\alpha_i\beta_j\eta)^{x_{ij}}\exp\left(-\alpha_i\beta_j\eta\right)\} \times \prod_{i}\left\{\alpha_i^{\delta-1}\exp\left(-\alpha_i\delta\right)\right\} \\ \times \prod_{j}\left\{\beta_j^{\delta-1}\exp\left(-\beta_j\delta\right)\right\} \times \eta^{\delta-1}\exp\left(-\eta\delta\right)\end{aligned} \tag{19}$$

and

$$\begin{aligned}\boldsymbol{\pi}(\alpha_j, \beta_i, \eta|\mathbf{x}) \propto \prod_{i\neq j}\{(\alpha_j\beta_i)^{y_{ij}}\exp\left(-\alpha_j\beta_i\right)\} \times \prod_{j}\left\{\alpha_j^{\delta-1}\exp\left(-\alpha_j\delta\right)\right\} \\ \times \prod_{i}\left\{\beta_i^{\delta-1}\exp\left(-\beta_i\delta\right)\right\}\end{aligned} \tag{20}$$

Thus, marginalising over each parameter in turn, we obtain the following conditional distributions;

$$\begin{aligned}\boldsymbol{\pi}(\alpha_i|\dots) &\propto \alpha_i^{\sum_{j\in O(i)} x_{ij}+\delta-1}\exp\left(-\alpha_i\left(\delta+\eta\sum_{j\in O(i)}\beta_j\right)\right) \\ \boldsymbol{\pi}(\beta_j|\dots) &\propto \beta_j^{\sum_{i\in O(j)} x_{ij}+\delta-1}\exp\left(-\beta_j\left(\delta+\eta\sum_{i\in O(j)}\alpha_i\right)\right) \\ \boldsymbol{\pi}(\eta|\dots) &\propto \eta^{\sum_{i,j} x_{ij}+\delta-1}\exp\left(-\eta\left(\delta+\sum_{i,j}\alpha_i\beta_j\right)\right)\end{aligned} \tag{21}$$

and

$$\begin{aligned}\boldsymbol{\pi}(\alpha_j|\dots) &\propto \alpha_j^{\sum_{i\in O(j)} y_{ij}+\delta-1}\exp\left(-\alpha_j\left(\delta+\sum_{i\in O(j)}\beta_i\right)\right) \\ \boldsymbol{\pi}(\beta_i|\dots) &\propto \beta_i^{\sum_{j\in O(i)} y_{ij}+\delta-1}\exp\left(-\beta_i\left(\delta+\sum_{j\in O(i)}\alpha_j\right)\right)\end{aligned} \tag{22}$$

Where in Equations 21 and 22, $O(i)$ represents the opponents of team $i$.
Thus, from these equations, we see that the conditional distributions of each parameter are of the forms;

$$
\begin{aligned}
\alpha_i | \ldots &\sim \text{Gamma}(\delta + \sum_{j \in O(i)} x_{ij}, \delta + \eta \sum_{j \in O(i)} \beta_j) \\
\beta_j | \ldots &\sim \text{Gamma}(\delta + \sum_{i \in O(j)} x_{ij}, \delta + \eta \sum_{i \in O(j)} \alpha_i) \\
\eta | \ldots &\sim \text{Gamma}(\delta + \sum_{i,j} x_{ij}, \delta + \sum_{i,j} \alpha_i \beta_j)
\end{aligned} \tag{23}
$$

and

$$
\begin{aligned}
\alpha_j | \ldots &\sim \text{Gamma}(\delta + \sum_{i \in O(j)} y_{ij}, \delta + \sum_{i \in O(j)} \beta_i) \\
\beta_i | \ldots &\sim \text{Gamma}(\delta + \sum_{j \in O(i)} y_{ij}, \delta + \sum_{j \in O(i)} \alpha_j)
\end{aligned} \tag{24}
$$

## 5.3 Process

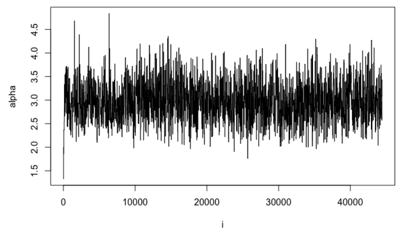
After deriving expressions for the conditional distributions expressed within Section 5.2, we proceed to implementing the corresponding Gibbs sampler algorithm.
This process begins with setting an initial (constraining) value for the $\delta$ parameter, thus enabling the induction of the algorithm and allowing parameter identifiability. Supposing we wish to obtain $k$ samples of each parameter, we denote the $i^{th}$ sample of each AS parameter by $\boldsymbol{\alpha}^{(i)} = (\alpha_1^{(i)}, \alpha_2^{(i)}, \ldots, \alpha_{20}^{(i)})$, and similarly for the DS parameters, $\boldsymbol{\beta}^{(i)}$. We must also implement initial values for some $\boldsymbol{\alpha}^{(i)}$ and $\boldsymbol{\beta}^{(i)}$, as well as implementing one for $\eta$ to begin the algorithm.
We found initial values of $\delta = 10$, $\boldsymbol{\alpha}^{(i)} = \boldsymbol{\beta}^{(i)} = 1$, and $\eta = 1.2$ to be adequate for the eventual convergence of the algorithm.

After setting initial values for the desired parameters, we update these by drawing from the aforementioned conditional distributions, updating and obtaining $\boldsymbol{\alpha}^{(i+1)}$, $\boldsymbol{\alpha}^{(i+2)}$, by sampling from the distribution of the previous iterations. This process is repeated until the parameter estimates have converged.

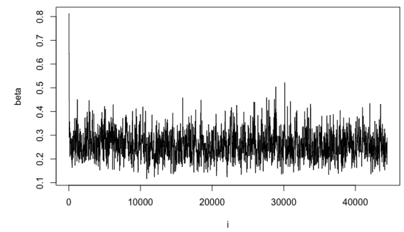
## 5.4 Fitting

### 5.4.1 Assessing Validity

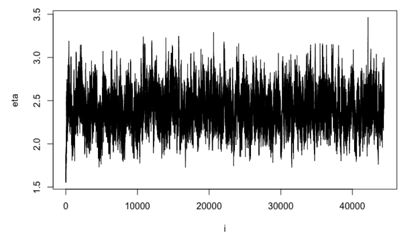
After applying the previously-described model to the data (we again use the, $8^{th}$ season of study, 2016/17), we must first assess whether the algorithm has converged sufficiently. This is done via the observation of trace-plots for each of the parameters. Within Figure 9 below, we observe representative traceplots of $\alpha_1$ and $\beta_1$, as well as the traceplot for $\eta$.




(a) $\alpha_1$




(b) $\beta_1$




(c) $\eta$

Figure 10: Traceplots of $\alpha_1$, $\beta_1$ and $\eta$ parameters.

From the respective $\alpha_1$, $\beta_1$ and $\eta$ Gibbs sampler traceplots within Figure 10, we see that there exists evidence of sufficient mixing within the iterative process, which is desirable. It would also appear from the plots that each of the processes have converged by the $40{,}000^{th}$ iteration, which is enough to conclude that the Gibbs sampler was satisfactory for obtaining the subsequent parameter estimates.

### 5.4.2 Parameter Estimates

In order to contrast the differing results of the Bayesian and frequentist approaches to modelling our data, we compare the parameter estimates obtained using the Gibbs sampling method to those obtained via a representative frequentist model.

The parameter estimates for each team are observed within Figure 11. Note that one immediate benefit of the Bayesian approach to modelling evident from Figure 11 is the constraint. Fixing the value of $\delta$ allows for the calculation of every team's standard error as opposed to having to setting one team's AS and DS parameters $= 1$ and comparing. This allows for a fuller analysis of team parameter comparisons, as each team has a posterior standard deviation upon their parameter estimates.

Within Section 4.2.2, we observed that the rankings of each team's parameter estimates

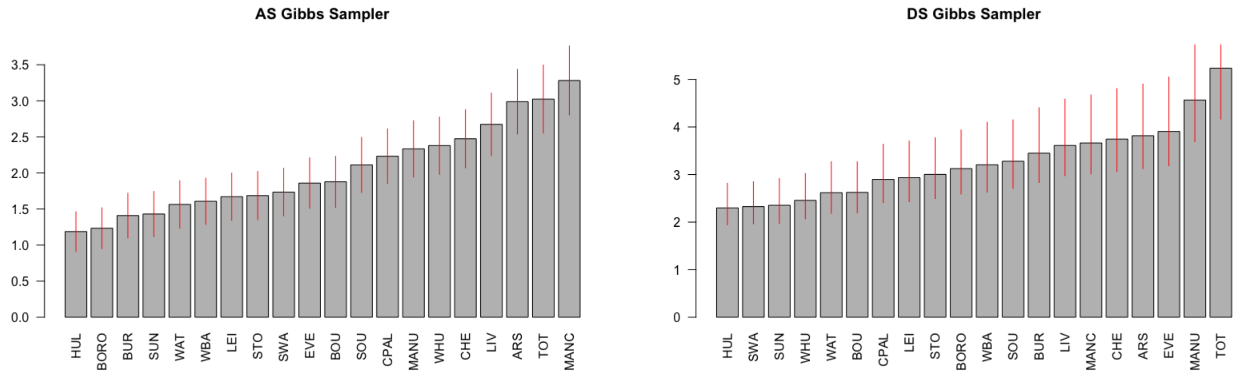



Figure 11: AS and DS for each team in the $8^{th}$ season of study obtained by Gibbs sampling. Red bars indicate posterior standard deviations of the parameter estimates.

seemed to sufficiently concur with the results of the 2016/17 season, with the AS parameter rankings appearing to reflect with the goals scored by each team throughout the season, and the DS parameter reflecting the number of goals against each team.
In order to assess the apparent accuracy of the Bayesian approach to modelling the data then, it suffices to compare the parameter rankings obtained by the two modelling approaches. To do this, we compare the estimates displayed within Figures 9 and 11 (Model 5 vs the Gibbs sampling estimates).

Observing the estimates found within Figure 11 alongside Figure 9 then, we see vast differences. For example, despite Tottenham having scored the most goals in the 2016/17 season, they are ranked $2^{nd}$ in terms of AS, with Manchester City (who scored the $3^{rd}$ most goals) ranked $1^{st}$ in terms of AS. Similar trends are apparent within the lower-reaches of the AS rankings, with Hull ranked $20^{th}$, even though they scored more goals within the 2016/17 season than both Middlesbrough (ranked $19^{th}$) and Sunderland (ranked $17^{th}$). Burnley (ranked $18^{th}$ in terms of AS), also seem to be out of place, having scored more goals than Sunderland within the season.

The DS parameters for the model seem slightly more reasonable, with Tottenham and Manchester United's DS ranks concurring with their 'goals conceded' within the season. However, Everton (ranked $3^{rd}$) seem to be significantly out of place. A similar positive trend though-can be observed within the lower half of the rankings, with the rankings generally agreeing with those from Model 5.

Despite the positive nature of the DS parameter rankings, this Bayesian model appears to be inadequate for modelling the data when observed alongside the frequentist approach found within Model 5.

# 6 The Sequential Bayesian Approach

It would seem from the previous section that the Bayesian approach to modelling our data via the multiplicative model has flaws. This Bayesian Gibbs sampler also requires a lot of computation to employ sequentially. This is due to the fact that each time a new observation is encountered, the algorithm must be run until convergence once more. In addition to this, it would appear that this approach does not provide significant improvement on the accuracy of the previously employed frequentist models.
Despite these conclusions however, there still exists the possibility of utilising an adapted Bayesian approach effectively to model the data for the season.

The initial Bayesian approach's inefficiency in part stemmed from its inability to incorporate new observations without the need for the algorithm to be run to completion. This results in the MCMC taking a significant amount of time to implement. Thus, instead of using the full seasons' data to update the priors defined at the outset of each season, it may be more desirable and productive to utilise the nature of the Bayesian approach within a sequential framework.
As before, this approach hinges upon the setting of priors at the outset of the season. However, as opposed to updating the corresponding posteriors after the final round of the season, we instead update them round by round throughout the season (after each set of matches), accounting for each round's results within the parameter estimates for each team.

A sequential Bayesian approach to modelling the seasons' data allows us to observe the development of each team's attacking and defensive abilities throughout the season, giving us a more in-depth illustration of the data within the season. The formulation of this model is slightly more complex, but convergence time is not an issue within this framework.

## 6.1 Defining Terms

As was the case in preceding sections, within the sequential Bayesian framework, we again set $i \in \{1,2,\ldots,n\}$ to be the home team within each match, and $j \in \{1,2,\ldots,n\}$ to be the away team. Again, the value of $n$ is the number of teams within the league in question (for the English Premier League, $n = 20$).
Within this model, we update parameter estimates based upon each round's results. As this is the case, we define an indicator for the round number: $R(r)$, where $r \in \{1,2,\ldots,38\}$. Now we may let $\{X_{ij}, Y_{ij}\}^{(r)}$ represent the respective score of home team $i$ and away team $j$ in each round $R = r$ of study.

We formulate the models for each parameter as below;

$$\begin{gathered} X_{ij} \sim \text{Poisson}(\lambda_{ij}^{(r)}) \quad \& \quad Y_{ij} \sim \text{Poisson}(\mu_{ij}^{(r)}) \\ \text{With:} \quad \lambda_{ij}^{(r)} = \alpha_i^{(r)} \beta_j^{(r)} \eta^{(r)} \quad \& \quad \mu_{ij}^{(r)} = \alpha_j^{(r)} \beta_i^{(r)} \end{gathered} \tag{25}$$

Within Equation (25) as before, $\alpha_k$ and $\beta_k$ represent the AS and DS of team $k$ respectively, and $\eta$ the common HGA.

## 6.2 Finding Expressions for Conditional Distributions

In order to find the used conditional distributions, we must first define initial priors for each of the parameters utilised within the model. These are similar to those used within the initial Bayesian approach considered within S5.2, and are outlined below;

$$\begin{aligned} \alpha_i^{(r)} &\sim \text{Gamma}(\nu_{a,i}, \nu_{b,i}) \\ \beta_i^{(r)} &\sim \text{Gamma}(\xi_{a,i}, \xi_{b,i}) \\ \eta^{(r)} &\sim \text{Gamma}(\kappa_a, \kappa_b) \end{aligned} \tag{26}$$

The distribution functions within Equation (25) for each parameter lead to the likelihood found within Equation 27 (below) for round $R = r$;

$$\mathcal{L}(\boldsymbol{\alpha}, \boldsymbol{\beta}, \eta) \propto \prod_{\{i,j\} \in R(r)} \exp\left(-\{\alpha_i \beta_j \eta + \alpha_j \beta_i\}\right) \times \{\alpha_i \beta_j \eta\}^{x_{ij}^{(r)}} \times \{\alpha_j \beta_i\}^{y_{ij}^{(r)}} \tag{27}$$

This likelihood, combined with the prior distributions found within Equation (26), lead to the joint posterior distribution after round $R = r$ as below;

$$\begin{aligned} \boldsymbol{\pi}(\boldsymbol{\alpha}, \boldsymbol{\beta}, \eta | \mathbf{x}^{(r)}, \mathbf{y}^{(r)}) \propto{}& \exp\left(-\sum_{\{i,j\} \in R(r)} \{\alpha_i \beta_j \eta + \alpha_j \beta_i\}\right) \times \prod_{\{i,j\} \in R(r)} \{(\alpha_i \beta_j \eta)^{x_{ij}^{(r)}} \times (\alpha_j \beta_i)^{y_{ij}^{(r)}}\} \\ &\times \left[\prod_{i=1}^{n} \alpha_i^{\nu_{a,i}^{(r)}-1} \exp(-\alpha_i \nu_{b,i}^{(r)})\right] \times \left[\prod_{j=1}^{n} \beta_j^{\xi_{a,j}^{(r)}-1} \exp(-\beta_j \xi_{b,j}^{(r)})\right] \times \eta^{\kappa_a^{(r)}-1} \exp(-\eta \kappa_b^{(r)}) \end{aligned} \tag{28}$$

As was the case within the non-sequential Bayesian section, we reformulate this joint posterior to produce the full conditionals for each of the parameters within the model, $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$, $\eta$. These are found simply again by marginalising over the parameter in question, and the formulation for these distributions are found below;

$$
\begin{aligned}
\boldsymbol{\pi}(\alpha_i|\dots) &\propto \alpha_i^{\nu_{a,i}^{(r)}+\sum_{j\in O(i)} x_{ij}^{(r)}-1} \exp\left(-\alpha_i\{\nu_{b,i}^{(r)}+\eta\sum_{j\in O(i)}\beta_j\}\right) \\
\boldsymbol{\pi}(\alpha_j|\dots) &\propto \alpha_i^{\nu_{a,j}^{(r)}+\sum_{i\in O(j)} y_{ij}^{(r)}-1} \exp\left(-\alpha_j\{\nu_{b,j}^{(r)}+\sum_{i\in O(j)}\beta_i\}\right) \\
\boldsymbol{\pi}(\beta_j|\dots) &\propto \beta_j^{\xi_{a,j}^{(r)}+\sum_{i\in O(j)} x_{ij}^{(r)}-1} \exp\left(-\beta_j\{\xi_{b,j}^{(r)}+\eta\sum_{i\in O(j)}\alpha_i\}\right) \\
\boldsymbol{\pi}(\beta_i|\dots) &\propto \beta_i^{\xi_{a,i}^{(r)}+\sum_{j\in O(i)} y_{ij}^{(r)}-1} \exp\left(-\beta_i\{\xi_{b,i}^{(r)}+\sum_{j\in O(i)}\alpha_j\}\right) \\
\boldsymbol{\pi}(\eta|\dots) &\propto \eta^{\kappa_a^{(r)}+\sum_{i,j} x_{ij}^{(r)}-1} \exp\left(-\eta\{\kappa_b^{(r)}+\sum_{i,j}\alpha_i\beta_j\}\right).
\end{aligned}
\tag{29}
$$

Within Equations (29), $O(i)$ is defined similarly to in Section 5.2, denoting the set of opponents of team $i$.

In order to obtain parameter estimates with respect to these conditional posterior distributions, we may pursue approximations of the sufficient statistics using modal values from a converged MCMC. These sufficient statistics are listed below and pertain to estimates for each round after round $R = r$.

$$
\begin{array}{lll}
\nu_{a,i}^{(r+1)} = \nu_{a,i}^{(r)} + x_{ij}^{(r)} & \nu_{b,i}^{(r+1)} = \nu_{b,i}^{(r)} + \tilde{\eta}\tilde{\beta}_j & \textbf{HOME} \\
\nu_{a,j}^{(r+1)} = \nu_{a,j}^{(r)} + y_{ij}^{(r)} & \nu_{b,j}^{(r+1)} = \nu_{b,j}^{(r)} + \tilde{\beta}_i & \textbf{AWAY} \\
\xi_{a,i}^{(r+1)} = \xi_{a,i}^{r} + y_{ij}^{(r)} & \xi_{b,i}^{(r+1)} = \xi_{b,i}^{(r)} + \tilde{\eta}\tilde{\alpha}_j & \textbf{HOME} \\
\xi_{a,j}^{(r+1)} = \xi_{a,j}^{r} + x_{ij}^{(r)} & \xi_{b,j}^{(r+1)} = \xi_{b,j}^{(r)} + \tilde{\alpha}_i & \textbf{AWAY} \\
\kappa_a^{(r+1)} = \kappa_a^{(r)} + \sum_{i,j} x_{ij} & \kappa_b^{(r+1)} = \kappa_b^{(r)} + \sum_{i,j}\tilde{\alpha}_i\tilde{\beta}_j &
\end{array}
\tag{30}
$$

We aim to update each of the AS and DS parameters by way of updating the sufficient statistic (as shown above). However, we do not possess said sufficient statistic for the scale parameter within the Gamma distribution, as the values of $\alpha$, $\beta$ and $\eta$ are unknown.
We therefore approximate the Gamma scale parameter using the expectations from the previous rounds.

## 6.3 Fitting

### 6.3.1 Final Parameter Estimates

In order to get an initial gauge on the performance of the sequential Bayesian approach, we may observe the parameter estimates after the conclusion of the $8^{th}$ season of study (2016/17), as we did for the previously-described frequentist and non-sequential Bayesian models.
We may contrast the rankings of the teams' estimates in conjunction with the final league standings, comparing the corroboration as we have done previously with the other models, giving us a measure of the model's effectiveness in context.
Within Figure 12, we see representations of the parameter estimates for each team following round 38 of the 2016/17 season.

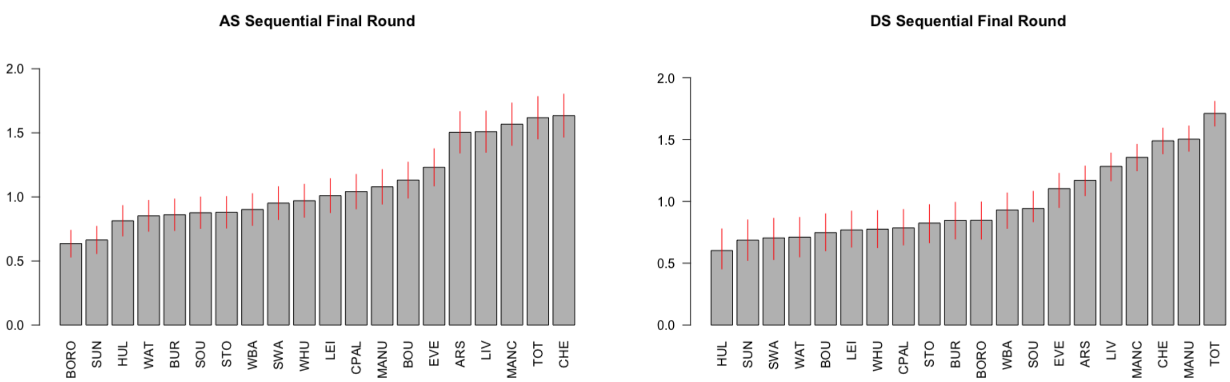


Figure 12: AS and DS for each team following the conclusion of the $8^{th}$ season of study obtained by the sequential Bayesian approach. Red bars indicate standard errors of the parameter estimates.

An initial positive we may take from Figure 12 is observed within the measures of the standard errors for each parameter, as each estimate is reliant on less samples than was the case in other models, giving more likely true representations of each estimate.

When we observe the rankings of each teams' parameter estimates in a similar fashion to the frequentist models, we see that the rankings of the estimates within this sequential model are identical to those produced by our initial frequentist model, Model 1. We deemed the parameter rankings produced within Model 1 to be sufficiently accurate in the context of the season's results, with this model only slightly out-performed by the Bivariate Dixon & Coles model, Model 5. Thus, the identical parameter rankings strongly indicate that the sequential Bayesian approach to modelling the Premier League data is *as suitable* as models reliant upon maximum likelihood estimation (the frequentist models) in terms of accuracy. This gives validation to our claims that a Bayesian approach would be suited to modelling this data, despite the Gibbs sampling approach being found to be inadequate in comparison to the frequentist approaches.

### 6.3.2 Sequential Benefits

After demonstrating that the sequential Bayesian approach to modelling the Premier League data, we proceed to illustrate the benefits of utilising this approach in comparison to the MCMC and maximum likelihood approaches.

Both the (MCMC) Gibbs sampling approach, and the maximum likelihood approach rely upon the full dataset for each season when constructing their respective parameter estimates. We have seen that utilising the full dataset produces a sufficiently accurate ranking of teams in terms of parameters within the maximum likelihood and (to a lesser extent) MCMC frameworks.
However, we have seen that updating parameter estimates sequentially after each round of the season results in rankings that are as accurate as those produced by the frequentist approach, and this sequential approach has the added benefit of illustrating the development of its parameter estimates throughout the duration of the season.

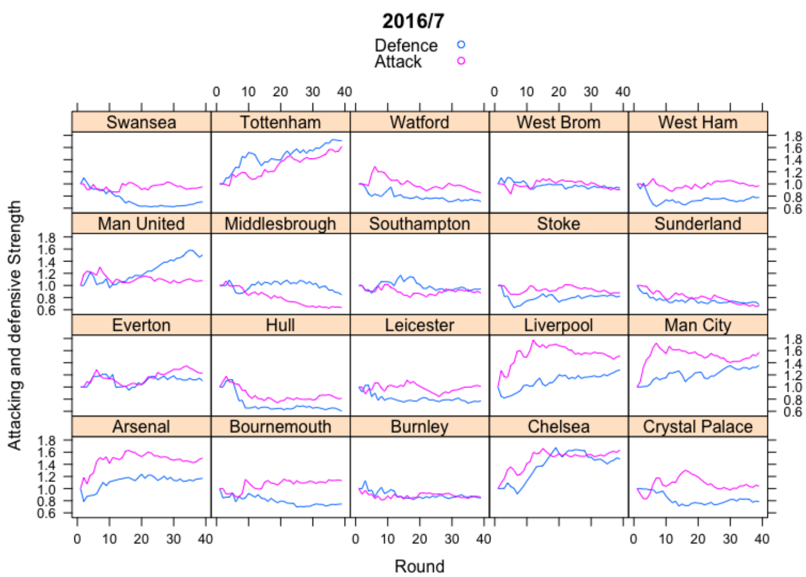



Figure 13: AS and DS development for each team in the $8^{th}$ season of study.

Within Figure 13, we observe the development of each team's AS and DS parameters over the course of the 2016/17 season.
These dynamic parameter representations help portray a more complete model of each teams' development over the course of the season. For example, despite finishing $5^{th}$ in the eventual DS rankings, Liverpool's early-season defensive form was similar to that of lesser teams in the league. Similar observations can be made with regards to eventual champions Chelsea. Chelsea's early-season defensive form was poor, despite being $3^{rd}$ in the final rankings behind Tottenham and Manchester United. Context can be provided for this upturn in form, with this coinciding with a change in defensive system (from a back four to a back three).
These observations are useful, as they give a fuller context to the eventual parameter estimates produced by the sequential model. This context is not available for the MCMC or

maximum likelihood approaches.

This approach also gives us real insight into how a consistent disparity between AS and DS parameters may affect a team's eventual ranking at the end of a season. For instance; the case of Manchester United. This team has a relatively low AS parameter when compared to the teams within the upper reaches of the final league table, despite finishing $6^{th}$. This is explained by the team's quickly established high DS parameter when compared to other similarly-ranked teams such as Liverpool and Arsenal for instance.

A similar effect may also be observed in the case of Bournemouth. Despite quickly establishing itself as a poor defensive side (as illustrated within Figure 12), its relatively high AS parameter over the course of the season prevent it from scoring low in terms of the eventual parameter estimate rankings.

These features may be observed to a lesser extent for the frequentist and MCMC approaches via the final parameter scores, but these illustrations of each teams' development give a fuller insight, and thus displays again the utility and convenience of the Bayesian sequential approach to modelling.

# 7 Prediction

We have found the Bayesian sequential approach to be quicker (and therefore simpler) to use than the MCMC Bayesian approach. We have also found this approach to be as accurate in modelling the Premier League data as the frequentist counterpart, whilst producing more useful and full information about its parameter estimates.

The aspect of the sequential modelling method which is of most practical use though, is its aptitude for forming predictions from its parameter estimates.

The Bayesian formulation of this approach allows for simpler construction of predictions (of match results) via the utilisation of the Skellam distribution. The frequentist maximum likelihood method does not allow for such simple calculations, as this would rely on the difficult computation of profile likelihoods, which is far from practical.

## 7.1 The Skellam Distribution

In order to form predictions of results from the parameter estimates developed via the Bayesian sequential method, we make use of the Skellam distribution.

The Skellam enables the modelling of the discrete probability distribution for a difference of two independent random variables. As described by [7] Karlis & Ntzoufras (2003), this distribution is applicable to the special case of dependent Poisson random variables in which each random variable has a common, random, 'covariance' contribution which is cancelled via the differencing of the two.

Within our analyses, we use the Skellam to model predictions for the difference between the number of goals scored by the home team, $i$ in a match between teams $i$ and $j$ at round R = r, $X_{ij}^{(r)}$, and the number of away goals scored within the same match, $Y_{ij}^{(r)}$. The

Skellam distribution produces expressions for these predicted differences. The formulation of the Skellam distribution is found within Equation 31, in which the terms are defined as in Subsections 6.1 and 6.2.

$$\mathbf{E}_{\boldsymbol{\theta}_{(r-1)}}(X_{ij}^{(r)}, Y_{ij}^{(r)}) \sim \text{Skellam}(\tilde{\eta}^{(r)}\tilde{\alpha_i}^{(r)}\tilde{\beta_j}^{(r)}, \tilde{\alpha_j}^{(r)}\tilde{\beta_i}^{(r)})$$
$$\text{Where :} \quad \tilde{\boldsymbol{\alpha}}^{(r)} = \frac{\boldsymbol{\nu_a}^{(r)}}{\boldsymbol{\nu_b}^{(r)}}, \quad \tilde{\boldsymbol{\beta}}^{(r)} = \frac{\boldsymbol{\xi_a}^{(r)}}{\boldsymbol{\xi_b}^{(r)}}, \quad \tilde{\eta} = \frac{\eta_a}{\eta_b} \tag{31}$$

As described by [7] Karlis & Ntzoufras (2003), if we set $K^r = X_{ij}^{(r)}$ - $Y_{ij}^{(r)}$, the difference between home and away goal counts, this distribution has the density listed within Equation 32.

$$\mathcal{P}\left(K^r = k\right) = \exp\left(\lambda_{ij}^{(r)} + \mu_{ij}^{(r)}\right)\left(\frac{\lambda_{ij}^{(r)}}{\mu_{ij}^{(r)}}\right)^{k/2} \mathcal{I}_k\left(2\sqrt{\lambda_{ij}^{(r)}\mu_{ij}^{(r)}}\right) \tag{32}$$

Where in Equation 32, $\mathcal{I}_k(x)$ is a modified Bezzel function with the formula;

$$\mathcal{I}_k(x) = \left(\frac{x}{2}\right)^{(k)} \sum_{r=0}^{\infty} \frac{\left(x^2/4\right)^r}{k!\Gamma(k+r+1)}$$

We use these score predictions within each matchup to estimate the probability of a home win, a draw, or an away win for each matchup. If we denote the result prediction for a game in the $r^{th}$ round as $\boldsymbol{\pi} = \{\pi_{HW},\, \pi_D,\, \pi_{AW}\}$, and we denote the outcome of this game as one of;

$$\mathbf{z}^{(r)} = \begin{cases} \{1\,0\,0\} & \text{Home Win} \\ \{0\,1\,0\} & \text{Draw} \\ \{0\,0\,1\} & \text{Away Win} \end{cases}$$

We calculate the Brier score in each round for our predictions as follows;

$$\mathbf{BS}_{(r)} = \left(\mathbf{z}^{(r)} - \boldsymbol{\pi}^{(r)}\right)^2$$

The closer this score is to 0, the more accurate the predictions we produce via the Bayesian sequential method are seen to be (as the predictions are closer to the real score for the seasons). In order to assess in context the adequacy of our model, we compare these scores to predictions made by other sources.

## 7.2 Comparison

As mentioned before, in order to ascertain whether the predictions made via our model are sufficiently accurate, we assess them using the Brier score. However, out of context we have no indication of the quality of these scores, thus we must make comparison to other sources.

Our choice of yard-stick to compare our model's Brier scores to lies within the odds produced by the betting company Bet365.

In order to obtain Brier scores from the odds published by Bet365 for each round within the season, we must first convert the odds into probabilities for use within the $\boldsymbol{\pi}^{(r)}$ parameters. We do this via the formula found within Equation 33.

$$\mathcal{P}(H) = \frac{1/O_H}{1/O_H + 1/O_D + 1/O_A} \tag{33}$$

Wherein $\mathcal{P}(H)$ is the supposed probability of a home win, $O_H$ the odds produced by Bet365 of a home win, $O_D$ similar for a draw, and $O_A$ the odds of an away win. For example if the odds produced by the company for a home win is 1/12, the odds of a draw 10/1 and the odds of an away win 20/1 (i.e. if the home team is a heavy favourite), then the corresponding probability of a home win ($\mathcal{P}(H)$) is $\frac{12/1}{12/11+1/10+1/20} = 0.357$. Analogous probabilities may be calculated for draws and away wins ($\mathcal{P}(D)$, $\mathcal{P}(A)$). In turn we use these to produce Brier scores for each round of the season.

## 7.3 Process

In order to carry out the process of Bayesian sequential predictions, we must adjust the aforementioned $\nu$, $\xi$ and $\kappa$ parameters outlined within S6.2. As well as updating throughout the process of a season, these updates also take into account the previous seasons' parameter estimates.
The updates are made to Equation 30, and can be found within Equation 34 with regards to a single observation, for simplicity.

$$\begin{array}{lll}
\nu_{a,i}^{(r+1)} \leftarrow \tilde{\nu}_{a,i}^{(r)} + x_{ij}^{(r)} & \nu_{b,i}^{(r+1)} \leftarrow \tilde{\nu}_{b,i}^{(r)} + \hat{\eta}^{r}\hat{\beta}_j^{\,r} & \textbf{HOME} \\
\nu_{a,j}^{(r+1)} \leftarrow \tilde{\nu}_{a,j}^{(r)} + y_{ij}^{(r)} & \nu_{b,j}^{(r+1)} \leftarrow \tilde{\nu}_{b,j}^{((r))} + \hat{\beta}_i^{\,(r)} & \textbf{AWAY} \\
\xi_{a,i}^{(r+1)} \leftarrow \tilde{\xi}_{a,i}^{r} + y_{ij}^{(r)} & \xi_{b,i}^{(r+1)} \leftarrow \tilde{\xi}_{b,i}^{(r)} + \hat{\eta}^{(r)}\hat{\alpha}_j^{\,(r)} & \textbf{HOME} \\
\xi_{a,j}^{(r+1)} \leftarrow \xi_{a,j}^{(r)} + x_{ij}^{(r)} & \xi_{b,j}^{(r+1)} \leftarrow \hat{\xi}_{b,j}^{(r)} + \hat{\alpha}_i^{\,(r)} & \textbf{AWAY} \\
\kappa_{a}^{(r+1)} \leftarrow \kappa_{a}^{(r)} + x_{ij}^{(r)} & \kappa_{b}^{(r+1)} \leftarrow \kappa_{b}^{(r)} + \hat{\alpha}_i^{\,(r)}\hat{\beta}_j^{\,(r)} & \textbf{prior HGA} \\
\kappa_{a}^{(r+1)} \leftarrow \kappa_{a}^{(r)} + x_{ij}^{(r)} + y_{ij}^{(r)} & \zeta_{b}^{(r+1)} \leftarrow \zeta_{b}^{(r)} + \hat{\alpha}_i^{\,(r)}\hat{\beta}_j^{\,(r)} & \textbf{HGA}
\end{array} \tag{34}$$

Thus, by applying the preceding updates, we are able to calculate parameter estimates for each team after each round of play as follows: $\hat{\boldsymbol{\alpha}}^{(r)} \leftarrow \frac{\boldsymbol{\nu_a}^{(r+1)}}{\boldsymbol{\nu_b}^{(r+1)}}$, $\hat{\boldsymbol{\beta}}^{(r)} \leftarrow \frac{\boldsymbol{\xi_a}^{(r+1)}}{\boldsymbol{\xi_b}^{(r+1)}}$ and $\hat{\eta}^{(r)} \leftarrow \frac{\kappa_a^{(r+1)}}{\kappa_b^{(r+1)}}$.

## 7.4 Brier Scores

Following the process described within S7.3, we produce Brier scores for each year, and assess in what proportion of the respective season's games our model's Brier scores were superior to those of Bet365. We calculate the Bet365 scores using the method within sections 7.1 and 7.2.

Being consistent with the previous sections of the paper, we assess the scores obtained for the 2016/17 season (the $8^{th}$ season of study), incorporating the 2015/16 season's final parameter estimates into the priors for this model.
A comparison of Brier scores for the model and Bet365's odds are observed within Figure 14.

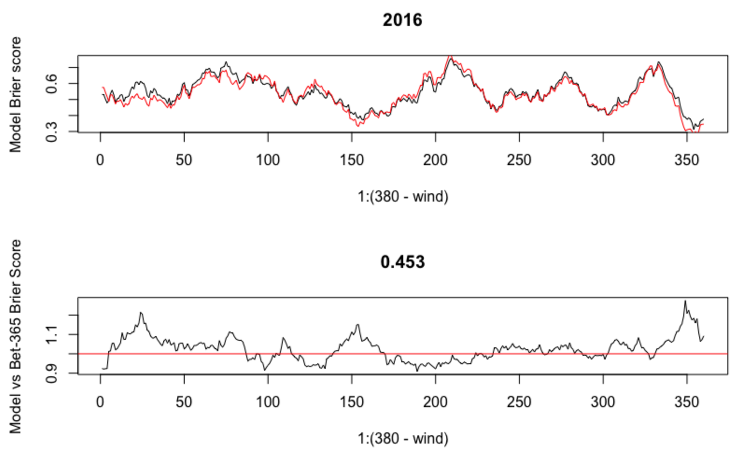



Figure 14: A Brier scores for the 2016/17 season obtained by model viewed alongside those obtained by Bet365's odds(top), Bet365 Brier score - Model Brier score with proportion of games for which the model is better as the heading (bottom)

We see from Figure 13 that our Bayesian sequential model appears to produce similar Brier scores to those obtained via conversion of Bet365's odds, with our model even outperforming Bet365 within 45.3% of the 2016/17 season's matches. We observe this comparison within all 10 years of study within Table 9.

Table 9 displays the adequacy of the predictions made by our Bayesian sequential model when compared to predictions obtained via conversions of the odds published by Bet365 for each game. In 2011/12 (highlighted in blue) especially, we see that the model outperformed Bet365's predictions for a slim majority of the games (52.6%) but this is enough to show the power of the Bayesian sequential approach when making predictions with regards to Premier League football results.

Table 13: Percentage of Bayesian Sequential model's Brier scores which are better than those obtained by conversion of Bet365 odds for each season of study

| Season | MBS >Bet365BS (%) |
|---|---|
| 2009/10 | 44.2 |
| 2010/11 | 48.9 |
| 2011/12 | 52.6 |
| 2012/13 | 39.5 |
| 2013/14 | 36.6 |
| 2014/15 | 40.5 |
| 2015/16 | 42.6 |
| 2016/17 | 45.3 |
| 2017/18 | 41.6 |

# 8 Concluding Remarks

## 8.1 Summary

Throughout this paper we have highlighted and developed numerous methods for modelling Premier League football scores. Initially, we displayed the fact that univariate Poisson models are adequate for the modelling of results, and subsequently the final league position of each team within a season. This reasserts the stance first put forward by [9] Maher, and then later by papers such as [7] Karlis & Ntzoufras that independent Poisson models (one for each of home and away goals counts) are as suited to modelling the scores of sports games as Negative Binomial models.
We compared and contrasted several univariate Poisson models, eventually concluding (on the basis of AIC and BIC scores, as well as the assessing of modelling assumptions) that a model possessing parameters which accounted for the Attacking Strength (AS) and Defensive Strength (DS) of each team, as well as accounting for a common Home Ground Advantage (HGA) was the most suited to modelling the data in question. We assessed this model for overfitting and concluded there was no evidence of this.

After obtaining the optimal univariate Poisson model for the data, we moved to investigating the assumption of independence between the supposed independent Poisson distributions of the home and away goal tallies within each game. In doing this we found there to be evidence enough of some correlation between home and away goal counts within games, which led us to utilise and investigate the possibility of Bivariate Poisson models put forward by [7] Karlis & Ntzoufras.
This Bivariate model though was not found to be significantly superior to the univariate model mentioned earlier, and it was thus concluded that accounting for covariance between home and away goal counts does not improve model fit.

After establishing what we believed to be the best model for our data utilising maximum likelihood estimation, we then assessed how Bayesian approaches to modelling would perform.

The first of these approaches relied upon MCMC and Gibbs sampling using conditional distributions for parameters with a formulation concurrent with the optimal frequentist maximum likelihood model.
This model was found to be impractical due to its need of long run-time to garner a sufficient number of iterations to construct accurate parameter estimates.

The second of our Bayesian approaches relied upon the sequential updating of parameter estimates after each round of games within a season. This was found to be more suited than the MCMC method, as it was as accurate at modelling final league position as the frequentist model outlined earlier, and had a very short run time in comparison to the MCMC.
This approach also allowed for the viewing of the development of teams' parameters over the course of a season, and also allowed us to make simple predictions for each of the games within a season. These predictions were found to be comparable with those of the betting company Bet365 when assessed via comparison of Brier scores, thus indicating that the sequential approach to modelling is a strong approach, with many benefits not found within the frequentist models.

## 8.2 Implications and Possible Extensions

The results found within this paper may have many practical implications for the future. With regard to prediction, the sequential Bayesian method may have potential to be used in modelling football more accurately than the frequentist models posited by previous papers such as [5], etc. This model may also be applied to other ball games such as hockey or outside of the realm of sports modelling all together.

Given unlimited time, I would have liked to have developed the Bayesian sequential approach (which I believe to have the most modelling potential), possibly via the inclusion of a forgetting function, in which the effects of previous parameter estimates are lessened over time, which I feel would be more realistic. Incorporating a correlation function into my sequential Bayesian model may have also improved my model, which is another thing I would have liked to have investigated given time.
I would also have liked to have investigated the model when applied to other similar ball games, an issue which shall be left for the reader's consideration.

# References


[1] Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, 19(6):716–722.

[2] Cattelan, M., Varin, C., and Firth, D. (2013). Dynamic bradley-terry modelling of sports tournaments. *Journal of the Royal Statistical Society*, 62(1):135–150.

[3] Clarke, S. and Norman, J. (1995). Focus on sport: Home ground advantage of individual clubs in english soccer. *Journal of the Royal Statistical Society*, 44(4):509–521.

[4] Cook, D. (1977). Detection of influential observation in linear regression. *Technometrics*, 19(1):15–18.

[5] Dixon, M. and Coles, S. (1997). Modelling association football scores and inefficiencies in the football betting market. *Journal of the Royal Statistical Society*, 46(2):265–280.

[6] Dobson, A. (2002). *An Introduction to Generalized Linear Models Second Edition*. Chapman Hall, London.

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

[8] Lee, A. (1997). Modelling scores in the premier league: Is manchester united really the best? *CHANCE*, 10(1):15–19.

[9] Maher, M. (1982). Modelling association football scores. *Statistica Neerlandica*, 36(3):109–118.

[10] Moroney (1951). *Facts From Figures*. Pelican, London.

[11] Reep, C., Pollard, R., and Benjamin, B. (1971). Skill and chance in ball games. *Journal of the Royal Statistical Society*, 134(4):623–629.

[12] Schwarz, G. (1978). Estimating the dimension of a model. *The Annals of Statistics*, 6(2):461–464.