Mathematics & Statistics | Lancaster University

Bayesian Dynamic Modelling of Football Results

Using a Bayesian approach to state space models for predicting football results

by

Matthew Kenneth Williams

Supervisor:

Gareth Ridall

Dissertation submitted in partial fulfilment for the degree of *Master of Science in Data Science*

September 2020

Contents

1 Introduction **2**
- 1.1 Motivation of the Dissertation 2
- 1.2 Objectives 3
- 1.3 Structure 3

2 Literature Review **5**

3 Static Models for Football **7**
- 3.1 Models for Comparison 8
  - 3.1.1 Rao Kupper Model 9
  - 3.1.2 Davidson Model 9
  - 3.1.3 Glickman Model 10
  - 3.1.4 Univariate Poisson Model 1 10
  - 3.1.5 Univariate Poisson Model 2 10
  - 3.1.6 Bivariate Poisson Model 11
  - 3.1.7 Bivariate Negative Binomial Model 11
  - 3.1.8 Dixon and Coles Model 11
  - 3.1.9 Ordinal Model 12
  - 3.1.10 Firth Model 12
  - 3.1.11 Bayesian Sequential Model 12
- 3.2 Methodology 13
- 3.3 Scoring Methods 13
  - 3.3.1 Brier Score 14
  - 3.3.2 Log Score 14
  - 3.3.3 Ranked Probability Score 15
- 3.4 Results 15
- 3.5 Summary 17

4 Bayesian State Space Models **19**
- 4.1 Motivation 19
- 4.2 Simple Bayesian State Space Model Exploiting Conjugacy 19
- 4.3 Summary 22

5 Bayesian State Space Football Models **23**

5.1 Model Explanation 23
5.2 Parameter Estimations 27
5.3 Models Created 28
5.3.1 Model 1 - Constant Within Season Forgetting Factor 29
5.3.2 Model 2 - Model Averaging for the Within Season Forgetting Factor 30
5.3.3 Method Results 31
5.4 Strengths Plots 33
5.5 Summary 35

6 Conclusion 38

Appendices 39
A Bayesian Dynamic Modelling of Football Results: Project Specification . . 40

Bibliography 43

List of Tables

5.1 Table of nuisance parameters optimised using the ten seasons of the Premier League from 1995/96 to 2004/05, for a model with a single within season forgetting factor 30
5.2 Table of nuisance parameters optimised using the ten seasons of the Premier League from 1995/96 to 2004/05, for a model using a model averaging approach for the within season forgetting factor 31

List of Figures

3.1 Average Brier scores for each optimised static model for each season, from 1995/96 to 2019/20. 16
3.3 Average log scores for each optimised static model for each season, from 1995/96 to 2019/20. 16
3.2 Average ranked probability scores for each optimised static model for each season, from 1995/96 to 2019/20. 17
3.4 Ranked probability scores for each optimised static model for each season, from 1995/96 to 2019/20. The years shown are the years the seasons ended. 18

4.1 A simple Bayesian state space model that exploits conjugacy, with the diagram, measurement equation and state equation shown. 20

5.1 Dependence structure of the Bayesian dynamic football model, showing the prior distributions after the extend step. 25
5.2 Dependence structure of the Bayesian dynamic football model, showing the posterior distributions after the observations are made. 25
5.3 Evolution of the weights for each within season forgetting factor value used in the model averaging method, for the 2005/06 to 2019/20 seasons 32
5.4 The strengths of all teams from the 2006/07 season using a dynamic Bayesian state space model with a model averaging method for the within season forgetting factor 33

5.5 Ranked probability scores for the static, single within season forgetting factor and model averaging Bayesian models, for the 2005/06 to 2019/20 seasons 34
5.6 The strengths of Manchester United shown from 2005/06 to 2019/2020 using a dynamic Bayesian state space model using a single within season forgetting factor 35
5.7 The strengths of Liverpool shown from 2005/06 to 2019/2020 using a dynamic Bayesian state space model with a model averaging method for the within season forgetting factor 36

Abstract

This report assesses the suitability of a Bayesian dynamic model for predicting football results, using a conjugate state space method. Firstly, various static models are compared for predicting results, from which the Bayesian conjugate method is seen to perform best. The theory of state space models is then outlined for a simple Bayesian model, before a more complex one is discussed for football. One of the key features of the football model is the use of within season forgetting factors. Using a single value for this factor is considered, as well as a model averaging approach, by creating separate models. These two dynamic models are compared, alongside a static model, by looking at ranked probability scores. The two dynamic models are seen to perform the best, but with little between them. The attacking and defensive strengths using the dynamic models are then plotted and compared, using Liverpool and Manchester United as example teams, with some insights into how these teams have performed over the years as well.

Chapter 1

Introduction

1.1 Motivation of the Dissertation

In 2020 football was seen to be the most popular sport in the world, with people enjoying in playing and watching the game [39]. Many professional football matches, particularly in the top European divisions, in countries such as England, Germany and Spain, can be viewed on television in numerous countries around the world. These international audiences have grown substantially since the turn of the millennium, increasing the value of the leagues to multi billion pound levels due to the broadcasting rights and sponsorship fees paid [40].

In this dissertation, I focus on the top division in England, the Premier League. In the Premier League, there is always a 'home team' and an 'away team'. The home teams play at their chosen stadium for the match, believed to give them an advantage for the most part [30]. There are often many more fans of the home teams present at games, giving loud vocal support, as well as unique aspects of the environment compared to other stadia, such as the pitch dimensions, that could give the home side an edge. There are three possible outcomes of games, known as results. These results are determined by the scores of the games, whereby the score is the number of goals scored by each team. The scores are often given in the format 'A-B', where A is the number of 'home goals' and B is the number of 'away goals'. If the home team scores the most goals, then the result is a 'home win'. If the away team scores the most goals, then the result is an 'away win'. If both teams score the same number of goals, or neither team scores any goals at all, then the result is a 'draw'.

The interest in football across the globe is also evident in the betting industry. Football is the most popular sport for betting, creating a highly valuable industry, particularly in the Asian markets. At the start of the 2019/20 Premier League season, 17 of the 20 teams involved had a partnership with a betting brand [40]. The value of this market is one reason why predicting football results is of interest. There are many companies that operate to try and correctly predict football results on a professional level [41]. These companies may incorporate various information streams in a sophisticated manner, such as team line ups, weather and so forth, to generate their predictions. This dissertation focuses solely on using historical match scores to create predictions; however this could be of interest for its inclusion in a model using other variables also.

For this dissertation I use a Bayesian dynamic model to predict the outcomes of games. A dynamic model differs from the classical approach in that the parameters involved can change over time, whereas in a classical approach they remain fixed. This method should be better suited for football since the form of teams change over time. Form is the ability of a team at a given moment. The most notable example of recent years was Leicester City, who with many of the same players throughout, went from narrowly escaping relegation from the Premier League in 2015 to winning the title in 2016, when one bookmaker famously priced them at 5000 to 1 to win the league at the start of the season [14]. In the Bayesian dynamic model, I use a conjugate approach and update the parameters using a single step at each time point as part of a state space model. This approach allows the model to be run quickly, with no simulations necessary. Within season and between season forgetting factors are used, which mean that data from further in the past will affect the current strength parameters in the model less than data that is more recent.

By using a tool to predict football results, insight can often be gained into teams or games themselves as well, with analysis that can be taken away. This could be from the variables used, to show changes in form or style of play over time. This makes the models more useful than just being predictive tools and allows for an understanding of what is causing the results they give. A model may also be used to rank teams in terms of their abilities, similar to the ELO system used in Chess [27].

1.2 Objectives

The main of objective of this dissertation is to investigate how a Bayesian dynamic model can be applied to football data to predict the outcomes of games using historical results. I aim to explain the methodology involved and compare a static version to similar models that have been created. I then look to create and compare two different dynamic models, one which uses a single within season forgetting factor and one that uses a model averaging approach, alongside a static model. Lastly, I look to explore the strength parameters created from the two different dynamic models, including some insights into the performances of a couple of teams in the Premier League.

1.3 Structure

Firstly, I give a literature review on different models that have been used in the past for predicting football results. I then explore these models further in a chapter regarding static models, also known as classical models. This includes seeing how a static Bayesian model using conjugacy performs against the others, using Premier League scores. I then give the theory behind Bayesian dynamic state space models, with forgetting factors implemented. Following this I explain how a Bayesian dynamic football model can be created. I discuss the 'nuisance' parameters in the model, that need to be estimated before it can be run. This includes the within season forgetting factor. For this, I consider keeping it as a constant value. Then, I consider using a model averaging approach, whereby models using different values for the within season forgetting factor are averaged by the evidence in favour of each, to give an output. Models are created and ran for both approaches. Both dynamic models are compared alongside a static Bayesian model. I then look at attacking and defensive strength plots calculated from each of these two dynamic models, and explain why they appear as they do. Liverpool and Manchester United are used as example teams for the plots. Finally, I give a conclusion of the report, outlining the key points and findings to take away.

Chapter 2

Literature Review

There are various models that have been developed and extended upon over the years for predicting the outcomes of football games. There are three main ways that models have been developed to create the predictions. Firstly, there are models that predict the final scores of the games, giving the number of home goals and away goals in the matches. This is usually done by the Poisson distribution, which seems to fit goals in football well. From this, the results of the games can be determined. Although the Poisson distribution fits well, goals themselves are not seen to be truly Poisson distributed. Goals are not equally likely to occur at all times of a match, with goals generally more likely to be scored at the start and end of each half [19]. Other distributions have been suggested, such as in the the Weibull-count model proposed by Boshnakov et al in 2016 [15]. The second kind of predictive method uses the Skellam distribution [38], which models the difference in the home and away goals in order to determine the predicted results. Here, the home and away goals are assumed to be independently Poisson distributed. The final method comes from models which directly output the predicted results of games.

There are models created in the past that were related to preferences to treatments that can be applied to football results modelling as well. The Bradley Terry model [4] was developed in 1952 to model preferences for different medical treatments. The Rao Kupper model [33] extended upon the Bradley Terry model to allow for an expression of 'no preference' to treatments, equivalent to a draw in football, giving a model that can be used to give probabilities of the results of games using variables for the overall abilities of the home and away sides. The Davidson model also extended on the Bradley Terry model [7] in a different formulation, assuming that the probability of a draw is proportional to the geometric mean of the probability of a home win or away win. The Glickman model [18] again gave another adaption of the Bradley Terry model to output probabilities of match results, considering chess and tennis as sports it could be applied to. In 2013 a dynamic Bradley Terry model was proposed [6] allowing for the abilities of the teams to vary over time. This paper looked at basketball as well as football specifically, considering the results in Serie A, the top division in Italy. This method involved using Monte Carlo Markov Chain (MCMC) simulations [29].

In 1982 Maher first introduced using a bivariate distribution to model football scores using parameters for the attacking and defensive strengths of the teams in each game [22]. This was done using different parameters for each team when at home compared to away. Dixon and Coles [9] considered a similar model to Maher. They used the same attack and defense parameters for home and away for all games, but with the inclusion of a common 'home ground advantage' parameter across all games. A constant downweighting parameter was proposed, which remains the same across all games, to take data from the past into account less for the likelihood of observations. They also included a dependence parameter, as they felt that the goals in low scoring games were correlated, where neither team scored more than one goal. In 2003 a bivariate Poisson model [21] was developed by Karlis and Ntzoufras, which included modelling common goals in each game. These common goals are not related to the abilities of the sides. Other interesting considerations have been made for the interaction of the home and away goals scored in matches. In 2017 Boshnakov et all considered a bivariate Weibull count model, which led to positive betting returns [3].

A dynamic Bayesian approach for football was suggested in 2011 by Owen, whereby MCMC methods [29] were used to estimate a variance parameter which allowed the attacking and defensive strengths of teams to change over time [28]. A sequential model was first proposed in 2019 by Ridall et al [36]. This model is like the Dixon and Coles model, with the exclusion of the dependence parameter. This was adapapted to become the first method of its kind, by using a conjugate method to update the parameters using a single update at each stage as part of a dynamic state space model. The process involved is seen to be a very neat approach and less complex than converting Dixon and Coles into a dynamic model. The sequential Bayesian model will be the focus of this dissertation. A bivariate negative binomial model was also proposed by Ridall et al [36]. This uses a Poisson model to predict the home and away scores using attacking and defensive strengths, along with a shared random effect.

Chapter 3

Static Models for Football

In this chapter different static models are compared for modelling football results. In static models, parameters remain fixed and do not change over time. They are simpler and less computationally complex than dynamic models, whereby the parameters can change over time. Although the parameters in static models are fixed, the estimations of these parameters can change over time as more information becomes available. By comparing different models, it can be seen which methods perform best in the static form before moving onto the more advanced dynamic models.

To perform this analysis, I have used data for matches in the English Premier League from the 1996/97 to 2019/20 seasons from football-data.co.uk [11]. In each of these seasons there are 20 teams in the division. The teams in the division change each season, as three teams are relegated to the division below, the Championship, whilst three teams are promoted into the Premier League from the Championship. In each Premier League season, every team faces each other team twice. Once at home and once away. This means that there are 380 games within a season. I used only the following information for modelling:

- The dates of the games
- The names of home and away teams in the games
- The full time scores of the games (home and away goals)

When predicting the results for a particular season, only the previous results in that given season are used. The first ten rounds of games are not be considered as part of the model comparisons. This is because the predictions are not expected to be good for the models up to this point, as there would not have yet been enough data to sufficiently estimate the model parameters.

Goals greater than seven have been truncated in the data for estimating model parameters, meaning goals greater than seven are set to seven instead. This is so that the updates to the strength parameters are not overly influenced by high scoring games due to special circumstances. When teams score more than seven goals in football, there are often red cards or injuries for the opposition team involved that have had a significant benefit in their favour.

I first explain the different models that I will be comparing. These include models that output scores of the home and away sides, ones that use the Skellam distribution [38] to give the difference in the home and away goals, and others that output the final directly. All the models require Monte Carlo simulations [29] to estimate the parameters, except the Bayesian sequential model whereby the parameters are instead updated at each stage using conjugacy. I then look at three different scoring methods for assessing the predictions of each model.

3.1 Models for Comparison

In this section I will explain the key maths behind the different static models that I will be comparing. The backgrounds of these models were covered in the literature review chapter. Firstly, here are the key notations used across all the models. Some of the notations are specific to just one model, while others are used in multiple models.

- $t \in \{1, 2, ..., 380\}$. These are the games identifiers, where the games are arranged in chronological order from earliest to latest. Often several matches are played at the same time, and so in this case the matches are arranged by alphabetical order of the home team name.
- h_t. This denotes the home team in game t.
- a_t. This denotes the away team in game t.
- X_t. This denotes the home goals modelled in game t.
- Y_t. This denotes the away goals modelled in game t.
- x_t. This denotes the actual home goals in game t.
- y_t. This denotes the actual away goals in game t.
- V_t. This denotes the number of common goals modelled between the home and away sides in game t.
- λ_t^H. This denotes the expected home goals in game t.
- λ_t^A. This denotes the expected away goals in game t.
- u_i. This denotes the overall ability of team i.
- u_i^H. This denotes the overall ability of team i when playing at home.
- u_i^A. This denotes the overall ability of team i when playing away.
- α_i. This denotes the attacking strength of team i.
- β_i. This denotes the defensive strength of team i.
- γ. This denotes a home ground advantage parameter.

- ◇ z_t. This denotes the result of game t. The possible values are: [1,0,0] for a home win, [0,1,0] for a draw and [0,0,1] for an away win.

All the models require identifiability constraints. This ensures that there is only one solution of the maximum likelihood estimations of the parameters, therefore making them unique.

3.1.1 Rao Kupper Model

The Rao Kupper model [33] outputs the probabilities for the final results using abilities at home and away. These are

$$P(z_t = [1,0,0]) = \frac{u_{h_t}}{u_{h_t} + \sigma u_{a_t}} \quad \text{(Home win)}$$

$$P(z_t = [0,1,0]) = \frac{(\sigma^2 - 1)u_{h_t}u_{a_t}}{(u_{h_t} + \sigma u_{a_t})(u_{a_t} + \sigma u_{h_t})} \quad \text{(Draw)}$$

$$P(z_t = [0,0,1]) = \frac{u_{a_t}}{u_{a_t} + \sigma u_{h_t}} \quad \text{(Away win)}$$

where $\prod_{i=1}^{20} u_i = 1$ is the identifiability constraint and $\sigma > 1$. In this model the identifiability constraint is needed otherwise there would be an infinite number of solutions in each case for the strength parameters, u_i. This can be seen by multiplying each u_i parameter by an arbitrary integer x. In this instance, the equations become:

$$P(z_t = [1,0,0]) = \frac{u_{h_t}x}{u_{h_t}x + \sigma(u_{a_t}x)} = \frac{x}{x} \times \frac{u_{h_t}}{u_{h_t} + \sigma u_{a_t}}$$

$$P(z_t = [0,1,0]) = \frac{(\sigma^2 - 1)(u_{h_t}x)(u_{a_t}x)}{(u_{h_t}x + \sigma(u_{a_t}x))(u_{a_t}x + \sigma(u_{h_t}x))} = \frac{x^2}{x^2} \times \frac{(\sigma^2 - 1)u_{h_t}u_{a_t}}{(u_{h_t} + \sigma u_{a_t})(u_{a_t} + \sigma u_{h_t})}$$

$$P(z_t = [0,0,1]) = \frac{u_{a_t}x}{u_{a_t}x + \sigma(u_{h_t}x)} = \frac{x}{x} \times \frac{u_{a_t}}{u_{a_t} + \sigma u_{h_t}}$$

and so, the probabilities remain unchanged.

3.1.2 Davidson Model

Like the Rao Kupper model [33], the Davidson model [7] also outputs the probabilities for the final results using abilities at home and away. These are

$$P(z_t = [1,0,0]) = \frac{u_{h_t}}{u_{h_t} + u_{a_t} + \nu\sqrt{u_{h_t}u_{a_t}}} \quad \text{(Home win)}$$

$$P(z_t = [0,1,0]) = \frac{\nu\sqrt{u_{h_t}u_{a_t}}}{u_{h_t} + u_{a_t} + \nu\sqrt{u_{h_t}u_{a_t}}} \quad \text{(Draw)}$$

$$P(z_t = [0,0,1]) = \frac{u_{a_t}}{u_{h_t} + u_{a_t} + \nu\sqrt{u_{h_t}u_{a_t}}} \quad \text{(Away win)}$$

where $\prod_{i=1}^{20} u_i = 1$ is the identifiability constraint and $\nu > 0$.

3.1.3 Glickman Model

Like the Rao Kupper [33] and Davidson models [7], the Glickman model [18] also outputs the probabilities for the final results using abilities at home and away. These are

$$
\begin{aligned}
P(z_t = [1,0,0]) &= \frac{e^{u_{h_t}-u_{a_t}}}{1+e^{u_{h_t}-u_{a_t}+e^{0.5(u_{h_t}-u_{a_t})}}} \quad \text{(Home win)}\\
P(z_t = [0,1,0]) &= \frac{e^{0.5(u_{h_t}-u_{a_t})}}{1+e^{u_{h_t}-u_{a_t}+e^{0.5(u_{h_t}-u_{a_t})}}} \quad \text{(Home win)}\\
P(z_t = [0,0,1]) &= \frac{1}{1+e^{u_{h_t}-u_{a_t}+e^{0.5(u_{h_t}-u_{a_t})}}} \quad \text{(Home win)}
\end{aligned}
$$

where $\sum_{i=1}^{20} u_i = 0$ is the identifiability constraint.

3.1.4 Univariate Poisson Model 1

This version of the univariate Poisson model [8] uses the attacking and defensive strengths of teams, as well as a home ground advantage. The model calculates the predicted home and away goals for each game as

$$
\begin{aligned}
X_t|\lambda_t^H \sim \text{Poisson}(\lambda_t^H), &\quad \lambda_t^H = \alpha_{h_t}\beta_{a_t}\gamma\\
Y_t|\lambda_t^A \sim \text{Poisson}(\lambda_t^A), &\quad \lambda_t^A = \alpha_{a_t}\beta_{h_t}
\end{aligned}
$$

where $\prod_{i=1}^{20} \beta_i = 1$ is the identifiability constraint. The Skellam distribution [38] is then used to calculate the probabilities of the match results.

3.1.5 Univariate Poisson Model 2

This alternative version [36] of the univariate Poisson model 1 [8] is considered, using the home and away abilities of teams. The predicted home and away goals for each game are calculated as

$$
\begin{aligned}
X_t|\lambda_t^H \sim \text{Poisson}(\lambda_t^H), &\quad \lambda_t^H = \frac{u_{h_t}^H}{u_{a_t}^A}\\
Y_t|\lambda_t^A \sim \text{Poisson}(\lambda_t^A), &\quad \lambda_t^A = \frac{u_{a_t}^A}{u_{h_t}^H}
\end{aligned}
$$

where $\prod_{i=1}^{20} u_i^A = 1$ is the identifiability constraint. The Skellam distribution [38] is then used to calculate the probabilities of the match results.

3.1.6 Bivariate Poisson Model

The Bivariate Poisson model [21] models the number common goals for the home and away sides, and also the excess goals on top of that by the home and away sides. That is:

$$(X_t - V_t) \sim \text{Poisson}(\lambda_t^H)$$
$$(Y_t - V_t) \sim \text{Poisson}(\lambda_t^A)$$
$$(V_t) \sim \text{Poisson}(\lambda)\ V_t = 0, ..., \text{Min}(X_t, Y_t)$$

The equations for the lambdas are the same as for the first univariate Poisson model [8].

3.1.7 Bivariate Negative Binomial Model

The bivariate negative binomial model [36] is the same as the univariate Poisson model 1 [8], except a correlation variable is also included. This is done by

$$X_t|\lambda_t^H \sim \text{Poisson}(\lambda_t^H) \quad \lambda_t^H = \alpha_{h_t}\beta_{a_t}\gamma\epsilon_t$$
$$Y_t|\lambda_t^A \sim \text{Poisson}(\lambda_t^A) \quad \lambda_t^A = \alpha_{a_t}\beta_{h_t}\epsilon_t$$
$$\epsilon_t \sim \text{Gamma}(\kappa, \kappa)$$

where $\prod_{i=1}^{20} \beta_i = 1$ is the identifiability constraint and $\kappa > 0$. The likelihood marginalised over the random effect is

$$\begin{aligned} \text{f}(x_t, y_t, \epsilon_t) & \int_{\epsilon} \text{f}(x_t, y_t, \epsilon_t|\alpha, \beta, \gamma, \kappa)\text{p}(\epsilon_t|\kappa)\, d\epsilon_t \\ &= \frac{\Gamma(\kappa + x_t + y_t)}{\Gamma(\kappa)\Gamma(x_t+1)\Gamma(y_t+1)} p_t^{x_t} q_t^{y_t} (1 - p_t - q_t)^{\kappa} \end{aligned}$$

where $p_t = \frac{\lambda_t^H}{\kappa + \lambda_t^H + \lambda_t^A}$ and $q_t = \frac{\lambda_t^A}{\kappa + \lambda_t^H + \lambda_t^A}$. This is a bivariate negative binomial model and is able to explain over-dispersion and positive correlation.

3.1.8 Dixon and Coles Model

The static Dixon and Coles model [9] is the same as the univariate Poisson model 1 [8], except a correlation variable is also included, but only for low scoring games (where neither the home or away side score two goals or more). This is done by

$$\Pr(X_t = x_t, Y_t = y_t) = \tau_{\lambda,\mu}(x_t, y_T)\frac{\lambda^{x_t} e^{-\lambda}}{x_t!}\frac{\mu^{y_t} e^{-\mu}}{y_t!}$$

where

$$\lambda = \alpha_{h_t}\beta_{a_t}\gamma$$
$$\mu = \alpha_{a_t}\beta_{h_t}$$

$$\tau_{\lambda,\mu}(x_t, y_t) = \begin{cases} 1 - \lambda\mu\rho & \text{if } x_t = y_t = 0 \\ 1 + \lambda\rho & \text{if } x_t = 0,\, y_t = 1 \\ 1 + \mu\rho & \text{if } x_t = 1,\, y_t = 0 \\ 1 - rho & \text{if } x_t = y_t = 1 \\ 1 & \text{otherwise.} \end{cases}$$

$$\max(-1/\lambda, -1/\mu) \leq \rho \leq \min(1/\lambda\mu, 1)$$

3.1.9 Ordinal Model

The ordinal model used is from the "polr" function from the Mass package in R [35]. This method fits a logistic or probit regression model to an ordered factor response. The ordered factoring takes into account that a draw is closer to a home win than an away win is.

3.1.10 Firth Model

The static Firth [6] model calculates the probability of a home win, draw or away win. for each game using:

$$\text{pr}(R_t \leq r_t) = \frac{\exp\{\delta_{r_t} + \gamma + u_{h_t} - u_{a_t}\}}{1 + \exp\{\delta_{r_t} + \gamma + u_{h_t} - u_{a_t}\}}, \quad r_t \in \{0, 1, 2\}$$

where $-\inf < \delta_0 < \delta_1 < \delta_2 = \inf$. The identifiability constraints are $\delta_0 = -\delta$ and $\delta_1 = \delta$, with $\delta \geq 0$.

pr$(R_t \leq 0)$ is the probability that the result was an away win. pr$(R_t \leq 1)$ is the probability that the result was an away win or a draw. pr$(R_t \leq 2)$ is the probability that the result was a home win or a draw or an away win. From this the probability of a draw can be deduced from pr$(R_t \leq 1)-$pr$(R_t \leq 0)$ and the probability of a home win can be deduced from $1-$pr$(R_t \leq 1)$.

This is the static model described in the dynamic Bradley Terry model paper [6].

3.1.11 Bayesian Sequential Model

The Bayesian sequential model [36] gives the expected number of home and away goals for each game, modelled by

$$X_t|\lambda_t^H \sim \text{Poisson}(\lambda_t^H) \quad \lambda_t^H = \alpha_{h_t}\beta_{a_t}\gamma_t$$
$$Y_t|\lambda_t^A \sim \text{Poisson}(\lambda_t^A) \quad \lambda_t^A = \alpha_{a_t}\beta_{h_t}$$
$$\alpha_i, \beta_i, \gamma \sim \text{Gamma}(\delta, \delta) \quad i = 1, 2, ..., 20$$

This model is similar to the univariate Poisson model 1 [8], except it exploits conjugacy due to gamma distributions of the attacking and defensive strength variables. Identifiability is maintained in the model by setting the δ variable to a fixed value. For the purposes of setting up this model for comparison, I have set the δ variable to be equal to 10. Once the δ variable is set, there is no identifiability issue at later stages, as the conjugate method creates exact and unique updates at each stage. The posterior distributions for the parameters are

$$\alpha_{k,t} \sim \text{Gamma}(p^{\alpha}_{k,t}, q^{\alpha}_{k,t}), \quad k = 1, ..., 20$$
$$\beta_{k,t} \sim \text{Gamma}(p^{\beta}_{k,t}, q^{\beta}_{k,t}), \quad k = 1, ..., 20$$
$$\gamma_t \sim \text{Gamma}(p^{\gamma}_t, q^{\gamma}_t)$$

Using i to represent the home team for game t and j to represent the away team for game t, the updates made after each observation are

$$p^{\alpha}_{i,t} \leftarrow p^{\alpha}_{i,t-1} + x_t, \quad q^{\alpha}_{i,t} \leftarrow q^{\alpha}_{i,t-1} + \hat{\gamma}\hat{\beta_{j,t}}$$
$$p^{\alpha}_{j,t} \leftarrow p^{\alpha}_{j,t-1} + y_t, \quad q^{\alpha}_{j,t} \leftarrow q^{\alpha}_{j,t-1} + \hat{\beta_{i,t}}$$
$$p^{\beta}_{i,t} \leftarrow p^{\beta}_{i,t-1} + y_t, \quad q^{\beta}_{i,t} \leftarrow q^{\beta}_{i,t-1} + \hat{\alpha_{j,t}}$$
$$p^{\beta}_{j,t} \leftarrow p^{\beta}_{j,t-1} + x_t, \quad q^{\beta}_{j,t} \leftarrow q^{\beta}_{j,t-1} + \hat{\gamma}\hat{\alpha_{i,t}}$$
$$p^{\gamma}_t \leftarrow p^{\gamma}_{t-1} + x_t, \quad q^{\gamma}_t \leftarrow q^{\gamma}_{t-1} + \hat{\alpha_{i,t}}\hat{\beta_{j,t}}$$

where $\hat{\alpha_{i,t}} = \frac{p^{\alpha}_{i,t-1}}{q^{\alpha}_{i,t-1}}$, $\hat{\beta_{i,t}} = \frac{p^{\beta}_{i,t-1}}{q^{\beta}_{i,t-1}}$ and $\hat{\gamma}_t = \frac{p^{\gamma}_{t-1}}{q^{\gamma}_{t-1}}$. For all the teams not involves in game t, the hyper parameters of their attacking and defensive strengths remain unchanged.

3.2 Methodology

Each model was run over the 25 seasons of the Premier League from 1996/97 to 2019/20. The sequential nature of the Bayesian model was ran using a δ value of 10. The sequential updates were very fast and efficient. For all other models, the parameters used for each season at each point in time were optimised against the log likelihoods, using the previous scores observed in that particular season. Optim in R was used for this, with a BFGS method [34]. The non Bayesian models therefore took a long time to run, since the parameter optimisations must be re-run every time there is a new observation.

3.3 Scoring Methods

Scoring methods are useful for assessing the suitability and fit of the predictive models created and their ability to forecast ahead of time. These methods assess how close the predictions were to reality later observed. Traditional statistics is concerned mainly with the fit of models to the observed data, rather than out of sample predictions, as is the case here.

To compare models, I have used three different scoring methods which I will show. Lower scores for each of these methods indicates better predictions.

The notation $z_{j,t}$ represents whether or not the result for game t was j. The possible values for j are $[1, 0, 0]$ (Home win), $[0, 1, 0]$ (Draw) or $[0, 0, 1]$ (Away win). If $z_{j,t}$ is 1, it means that the result was j, or if it is 0 it means that this was not the result.

3.3.1 Brier Score

The Brier score [5] is the first scoring method used. This is calculated by

$$\text{BS} = \frac{1}{n}\sum_{t=1}^{n}\sum_{j=1}^{3}(z_{j,t} - P_{j,t})^2$$

The Brier score is a proper scoring rule. That is, the score is optimised when the true probabilities are used for the predictions, P, meaning $z_{j,t} = P_{j,t}$. This will give a Brier score of 0. This scoring method treats the three possible match results as categorical. Considering these two model outputs for a particular game

- Model A: Home win (0.2), Draw (0.3), Away win (0.5)
- Model B: Home win (0.2), Draw (0.5), Away win (0.3)

If the result for this game was a home win, the Brier score for models A and B will be the same due to results being treated as categorical. That is

- Brier Score for Model A $= (1 - 0.2)^2 + (0 - 0.3)^2 + (0 - 0.5)^2 = 0.98$
- Brier Score for Model B $= (1 - 0.2)^2 + (0 - 0.5)^2 + (0 - 0.3)^2 = 0.98$

3.3.2 Log Score

The log score [2] is the second scoring method used. The formula is seen in equation 3.1.

$$\text{LS} = -\frac{1}{n}\sum_{t=1}^{n}\sum_{j=1}^{3}[z_{j,t}\log(P_{j,t}) + (1 - z_{j,t})\log(1 - P_{j,t})] \tag{3.1}$$

The log score is also a proper scoring rule, like the Brier score. Match results are also treated as categorical variables, like they are for the Brier score.

3.3.3 Ranked Probability Score

The ranked probability score (RPS) [10] is the third scoring method used and is calculated using equation 3.2.

$$\text{RPS} = \frac{1}{2n}\sum_{t=1}^{n}\sum_{k=1}^{2}(\sum_{j=1}^{k}(z_{j,t} - P_{j,t}))^2 \tag{3.2}$$

The ranked probability score is also a proper scoring rule, like the Brier score and log score. However, unlike the Brier score, the possible match results are treated as ordinal rather than categorical. A draw is deemed as being closer to a home win than an away win is. Therefore, the data is ordered as home win, then draw, then away win. And so, using models A and B considered previously for a particular game, we see

- RPS for Model A $= (1 - 0.2)^2 + ((1 - 0.2) + (0 - 0.3))^2 = 0.89$
- RPS for Model B $= (1 - 0.2)^2 + ((1 - 0.2) + (0 - 0.5))^2 = 0.73$

showing that the RPS is lower for model B since it gives a higher probability of a draw than an away win. For this reason, the ranked probability score is often favoured when comparing prediction models for football results.

3.4 Results

In Figures 3.1, 3.3 and 3.2 the average metric for each model across all seasons can be seen, using the ranked probability score, Brier score and log score respectively. In each of these plots, the models are ordered by their averages from lowest (best) to highest (worst).

The plots for the ranked probability and Brier scores are seen to be very similar, both in terms of the heights of the bars and the order of the models. The Bayesian model is seen to have a significantly better score than all the other models too on both plots.

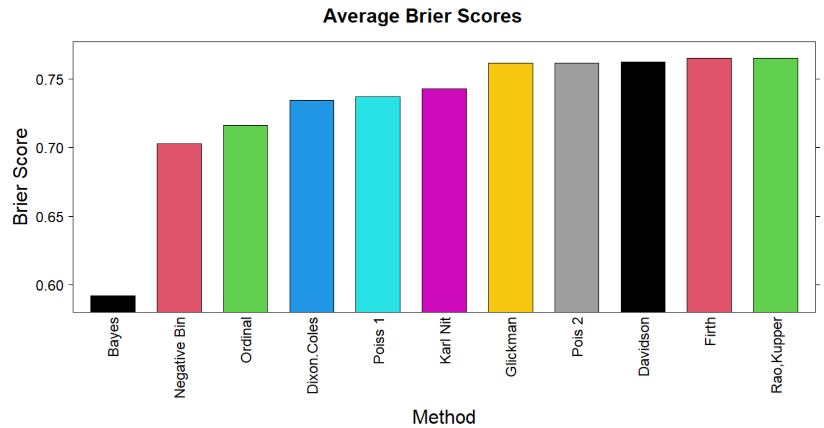

Figure 3.1: Average Brier scores for each optimised static model for each season, from 1995/96 to 2019/20.

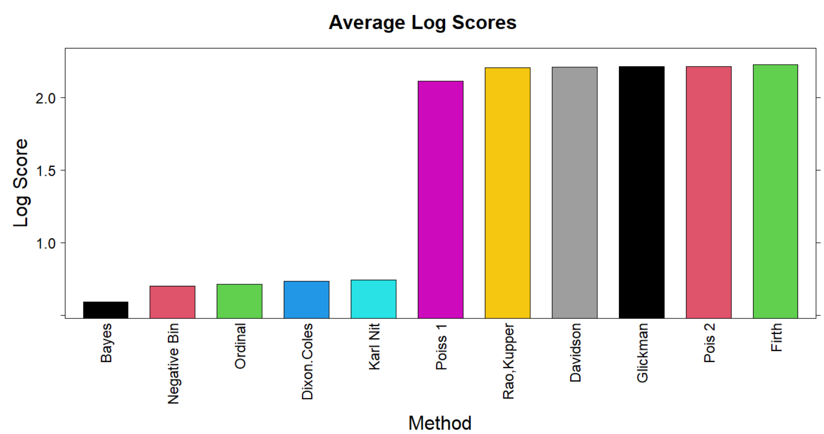

Figure 3.3: Average log scores for each optimised static model for each season, from 1995/96 to 2019/20.

The plot for the log scores is seen to be quite different than the ones for the ranked probability and Brier scores. More of a clear split is seen between the models, whereby the bars for the first five models are much lower than for the last six. This is because the log score penalises probabilities that are further away from their expected values more heavily. Although this is the case, the order of the bars is quite consistent. The three best models are ordered the same as for the ranked probability and Brier score plots. The Bayesian model is seen to be the best for this plot as well.

A Pearson correlation test was carried out between the three scoring methods. A high correlation of 0.98 was seen between the Brier scores and ranked probability scores. The log scores had a correlation of 0.61 and 0.59 with the Brier and ranked probability scores receptively. This is in line with the plots of the average scores seen using each

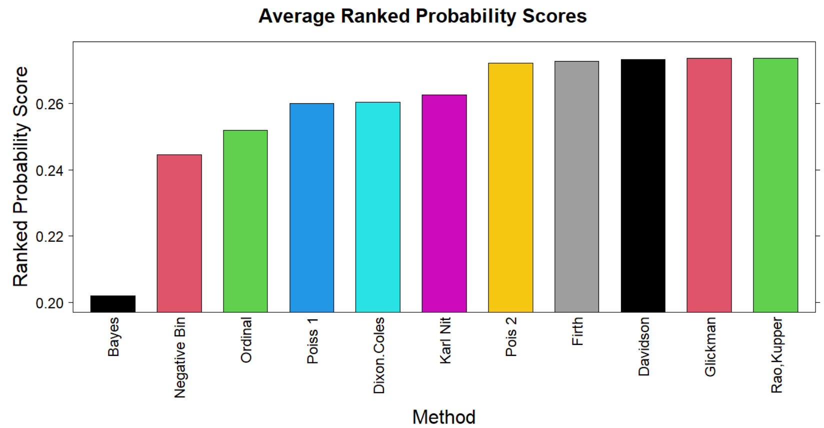

Figure 3.2: Average ranked probability scores for each optimised static model for each season, from 1995/96 to 2019/20.

method.

In Figure 3.4, the RPS is seen for each model across all seasons. The RPS is considered here since it is the most widely used metric for comparing football prediction models due to the fact it treats results as ordinal data. The bars are ordered in terms of the RPS model averages, seen in Figure 3.2, from best (left) to worst (right). A clear pattern can be seen from this plot across each of the seasons, whereby the height of the bars tends to increase from left to right in each season, showing that order of the averages seen are also generally observed in each individual season as well.

3.5 Summary

In this chapter, several different static models for predicting football results were explained and used to predict Premier League scores from 1996/97 to 2019/20. The average Brier scores, log scores and rank probability scores for each model were compared. The Bayesian sequential model was seen to have the best average for all three scoring metrics. The differences in the Brier and ranked probability scores between the models were seen to be very similar. The log score was seen to be somewhat different, with a clearer split between the models due to the greater penalisation this model gives for predictions further from the truth. These similarities and differences between the models were in line with correlation coefficients calculated. Lastly, the ranked probability score was observed for all models for each individual season. A clear trend was seen in line with the averages seen earlier, with the Bayesian model seen to consistently perform the best in every season.

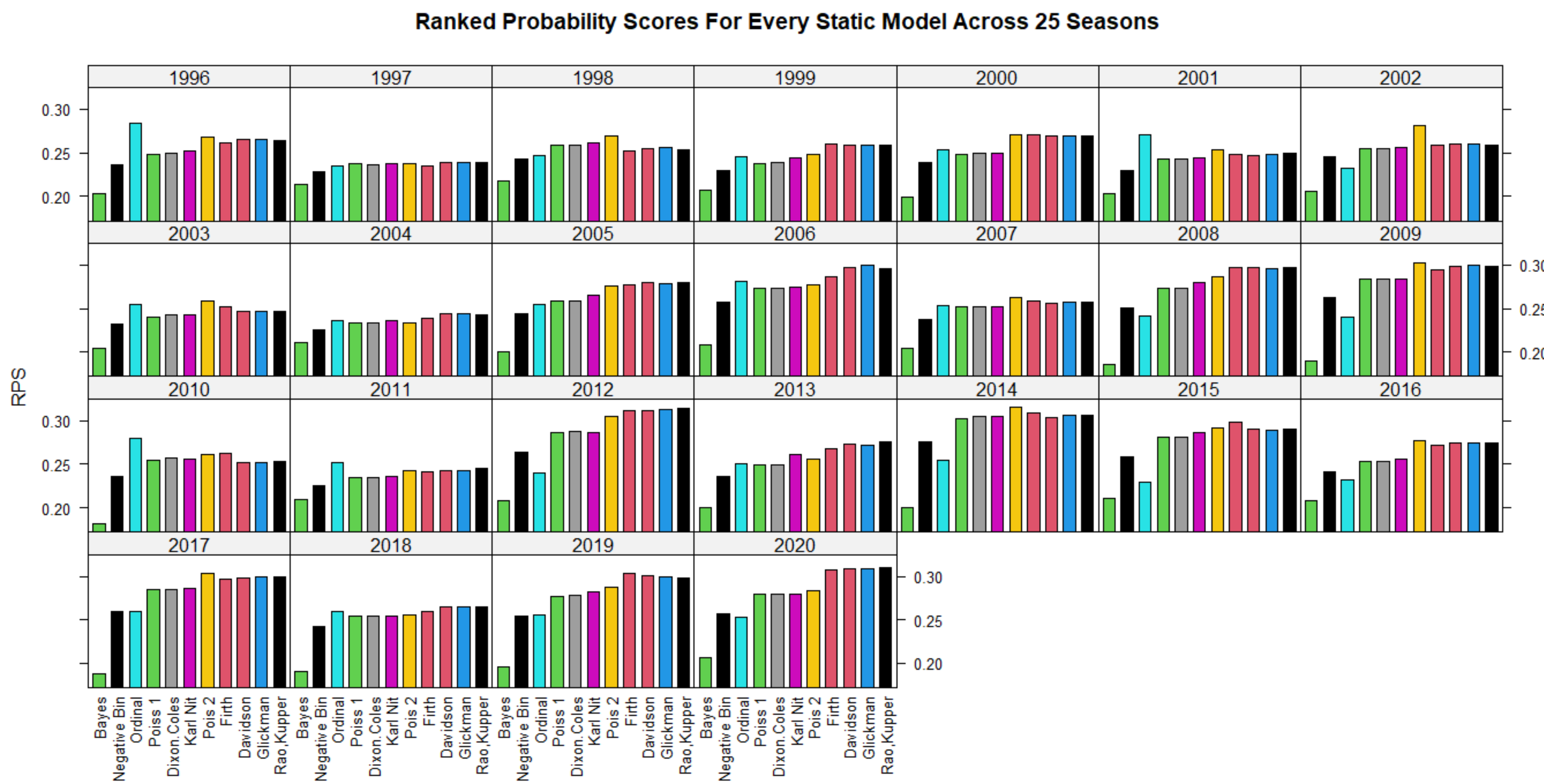

Figure 3.4: Ranked probability scores for each optimised static model for each season, from 1995/96 to 2019/20. The years shown are the years the seasons ended.

Chapter 4

Bayesian State Space Models

4.1 Motivation

A state space model was chosen for this project as it allows parameters to change over time. This is an important and useful feature for modelling in many scenarios. Examples include stock markets, where the volatility changes over time, and in pandemics where the rate of infection changes over time. This chapter explains Bayesian state space models that exploit conjugacy, the type that are later used on the football data.

In a Bayesian state space model, the parameters are modelled as random variables, that are 'updated' at each state from the previous state. Often in statistics, MCMC methods [29] are used to update parameters in a Bayesian model. However, this can be very time consuming and does not provide exact results. The alternative method of exploiting conjugacy provides exact closed form expressions to be used when calculating the parameters.

4.2 Simple Bayesian State Space Model Exploiting Conjugacy

In this section, a simple Bayesian state space model that exploits conjugacy is explained and discussed. This is seen visually in Figure 4.1. The observed values are denoted by y_t for each time point t. The measurements modelled, Y_t, are seen to be conditionally independent given θ_t, defined by Equation 4.1. The parameters θ_t are also seen to have the Markov property. This is defined by Equation 4.2.

$$Y_t|_{\theta_{1:t}} \perp Y_{t-1} \tag{4.1}$$

$$\pi(\theta_t|\theta_{1:t-1}) = \pi(\theta_t|\theta_{t-1}) \tag{4.2}$$

The likelihood of Y_t measurements are modelled by

$$\pi(Y_t|\theta_t) \sim \text{Poisson}(\theta_t) \quad \text{mean: } \theta_t \quad \text{variance: } \theta_t$$

The state equation links θ_{t-1} to θ_t, modelling the change in the parameter between states before new data is introduced. This is

$$\theta_t = \frac{Wt}{\omega_t}\theta_{t-1} \quad W_t \sim \text{Beta}(\omega_t p_{t-1}, (1-\omega_t)p_{t-1}) \quad 0 < \omega_t \leq 1$$

The mean of the Beta distribution for W_t is ω_t. Because of this, $E(\frac{W_t}{\omega_t}) = 1$, and so θ_t is modelled as following a random walk process. That is, for each change in state, θ_t is modelled as being equally likely to increase as it is to decrease. By exploiting conjugacy, the following gamma posterior for θ_t can be formed

$$\pi(\theta_t|y_{1:t}) \sim \text{Gamma}(p_t, q_t) \quad \text{mean: } \frac{p_t}{q_t} \quad \text{variance: } \frac{p_t}{q_t^2}$$

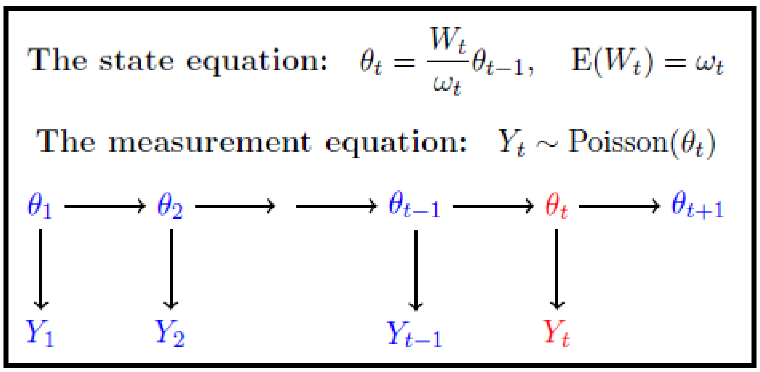

Figure 4.1: A simple Bayesian state space model that exploits conjugacy, with the diagram, measurement equation and state equation shown.

This conjugacy is created by using a prior for θ_t that is an extension of the posterior distribution at the previous state, $\pi(\theta_{t-1})$. The extension is created as

$$\overbrace{\pi(\theta_t|y_{1:t-1}, \omega_t)}^{\text{Extension}} = \int \overbrace{\pi(\theta_{t-1}|y_{1:t-1}, \omega_t)}^{\text{Posterior}} \overbrace{\pi(\theta_t|\theta_{t-1}, y_{1:t-1}, \omega_t)}^{\text{Transition}} \, d\theta_{t-1}$$
$$\propto \theta_t^{\omega_t p_{t-1} - 1} e^{-(\omega_t q_{t-1})\theta_t}$$

The proof can be seen in in [16]. This extended posterior distribution is seen to be $\text{Gamma}(\omega_t p_{t-1}, \omega_t q_{t-1})$. This is the same distribution as $\pi(\theta_{t-1})$, $\text{Gamma}(p_{t-1}, q_{t-1})$, but with both parameters multiplied by ω_t. These distributions have the same mean, $\frac{p_{t-1}}{q_{t-1}}$, except the variance of the extended posterior, $\frac{p_{t-1}}{\omega_t q_{t-1}^2}$, is larger. This is therefore said to be a 'forgetting' step in the model. As time t increases, observations further in the past have less of an effect on θ_t. The gamma prior for θ_t then creates a gamma posterior, due to the

Poisson likelihood for Y_t. This is shown by Equation 4.3.

$$\begin{aligned}
\overbrace{\pi(\theta_t|y_t)}^{\text{Posterior}} &\propto \overbrace{\pi(\theta_t)}^{\text{prior}} \times \overbrace{\pi(y_t|\theta_t)}^{\text{Likelihood}} \\
&\propto \theta_t^{\omega_t p_{t-1}-1} e^{-\omega_t q_{t-1}\theta_t} \times \theta_t^{y_t} e^{-n\theta_t} \\
&\propto \theta_t^{\omega_t p_{t-1}+y_t-1} e^{-(\omega_t q_{t-1}+n)\theta_t} \\
&\propto \theta_t^{p_t-1} e^{-q_t\theta_t}
\end{aligned} \tag{4.3}$$

The updates to the prior parameters in Equation 4.3 in order to form the posterior are

$$\begin{aligned}
p_t &\leftarrow \omega_t p_{t-1} + y_t \\
q_t &\leftarrow \omega_t q_{t-1} + n
\end{aligned}$$

One of the complications in this model is deciding which value to use for the forgetting parameter, ω_t. One approach is to run several models simultaneously, that each use different values for ω_t. The cumulative evidence for each model is then tracked. The value to use for ω_t is then calculated by weighting the ω_t values in the models that were run by their respective cumulative evidence values. To calculate the cumulative evidence, the predictive distribution is used. This is

$$\begin{aligned}
f(Y_t = y_t|y_{1:t-1}, \omega_t) &= \int \overbrace{\pi(\theta_t|y_{1:t-1,\omega_t})}^{\text{Extension}} \overbrace{f(y_t|\theta_t)}^{\text{Likelihood}} \, d\theta_t \\
&\propto \frac{1}{y_t!} \times \frac{\omega_t q_{t-1}{}^{\omega_t p_{t-1}}}{\Gamma(\omega_t p_{t-1})} \times \frac{\Gamma(p_t)}{q_t{}^{p_t}} \\
&\propto \frac{1}{y_t!} \times \frac{\Gamma(\omega_t p_{t-1} + y_t)}{\Gamma(\omega_t p_{t-1})} \times \left(\frac{1}{\omega_t q_{t-1} + 1}\right)^{y_t} \times \left(\frac{\omega_t q_{t-1}}{\omega_t q_{t-1}}\right)^{\omega_t p_{t-1}}
\end{aligned} \tag{4.4}$$

$$Y_t = y_t|y_{1:t-1}, \omega_t \sim \text{Negative-Binomial}\left(\omega_t p_{t-1}, \frac{\omega_t q_{t-1}}{\omega_t q_{t-1} + 1}\right)$$

The observed data at time t can be input into the predictive. If the predictive is high, it shows that the observed data was deemed likely and therefore matches well. This can then be used as 'evidence' for this value of ω_t at time t. The cumulative evidence at time t is then calculated by multiplying the evidence at each time from 2 to t. That is

$$Z_t(\omega) = f(y_{2:t}|y_1, \omega_t) = \prod_{t=2}^{t} f(y_t|y_{1:t-1}, \omega_t)$$

The first time used in the multiplication is 2, since this is the first time an extended prior, and therefore ω_t, is used in the model. If K models are used for model averaging, cumulative evidence for each ω_k value, for k in $\{1, 2, ..., K\}$ can be calculated as $Z(\omega_k)$. The model to use at time t to predict for time $t + 1$ can then be made using a mixture of all K models. This is

$$f(Y_{t+1}|y_{1:t}) = \sum_{k=1}^{K} \Omega_{k,t} f(Y_{t+1}|y_{1:t}, \omega_k)$$

where

$$\Omega_{k,t} = \frac{Z_{k,t}}{\sum_{k=1}^{K} Z_{k,t}}.$$

The posterior of θ_t can be modelled by

$$\pi(\theta_t|y_{1:t}) = \sum_{k=1}^{K} \Omega_{k,t} \pi(\theta_{k,t}|y_{1:t}, \omega_k) \tag{4.5}$$

with the expectation modelled as

$$\mathrm{E}(\theta_t|y_{1:t}) = \sum_{k=1}^{K} \Omega_{k,t} \mathrm{E}(\theta_{k,t}) = \sum_{k=1}^{K} \Omega_{k,t} \frac{p_{k,t}}{q_{k,t}}$$

An important feature of this model is that the evidence obtained from the predictive distribution, seen in equation 4.4, will be higher for lower forgetting factors when more unexpected results are observed. This is because unexpected results that are out of line with previous observations will not be predicted well with a model that places a lot of emphasis on those previous observations for the prior distributions. If a less informative prior is used by way of the forgetting parameter, this will lead to predictions that are more vague from the model, which will encapsulate more of these unexpected observations. Therefore if evidence is highest for a low forgetting factor, then this shows that the observed data is 'stable' with not too many unexpected observations, whereas if a high forgetting factor has the most evidence, then this shows that the observed data is 'unstable' with many unexpected observations present. It is also important to note that the evidence used for the model averaging approach may not lead to better predictions out of sample. Different models should therefore be run, with their performances assessed out of sample, before deciding which approach to take.

4.3 Summary

In this chapter, a Bayesian state space model was explained, which is a useful method for dynamic modelling. These models provide a clear process for how parameters change from state to state. The sequential Bayesian method shown gives single updates to the parameters at each stage by using a conjugacy. A model averaging approach was then explained, whereby the output is weighted from a range of models using different ω_t values. These models are then weighted against their cumulative evidences given the observations in order to provide the final output. High evidence for low forgetting factor parameters shows that the observed data is unstable, with many unexpected results present.

Chapter 5

Bayesian State Space Football Models

In football, the form of teams is seen to change over time. This is a popular topic of conversation amongst observers of the sport, as to why this may be or how it can be changed. Examples of why form may change include the players which are available, changes in tactics or the confidence levels of the players.

A dynamic model seems necessary for football, to encapsulate the changing form of teams over time. The Bayesian sequential model discussed in 3.1.11 can be adapted for this, by considering the attacking and defensive strengths of teams at each point in time using forgetting parameters, so data from further in the past is relied on less than data closer in time. This model then becomes a dynamic Bayesian state space model, which was described in a simpler form in the previous chapter using only one parameter θ_t. This method models the attacking and defensive strengths of team separately, for each point in time. The defensive and attacking strengths of teams do not always go hand in hand and so modelling the strengths separately allows for these differences to be observed and incorporated into the model.

Using a dynamic Bayesian model for football also allows the model to use data from past seasons in a reasonable way in the updating process. This can be done by including a forgetting parameter between seasons as well as within seasons. The within season forgetting parameter is used when creating the prior distributions for $t = 2$ to $t = 38$ for each season. The between season forgetting factor is used when creating the extended posterior after the final round of games in each season, to be used as the prior distributions for the start of the next season at $t = 1$.

In the rest of this chapter, I explain how the dynamic Bayesian sequential model works for football, before creating two different approaches. These approaches are one that uses a single within season forgetting factor, and another that uses model averaging. The results from these models will then be assessed and compared, alongside a static Bayesian sequential model

5.1 Model Explanation

In this section I will explain the main set up of the models being used. Firstly, all of the variables used for the dynamic football model are

- $t \in (1, 2, ..., 380)$. These are the games identifiers, where the games are arranged in chronological order from earliest to latest. Often several matches are played at the same time, and so in this case the matches are arranged by alphabetical order of the home team name.

- ◇ i, h_t. These denote the home team in game t.
- ◇ j, a_t. These denote the away team in game t.
- ◇ X_t. This denotes the home goals modelled in game t.
- ◇ Y_t. This denotes the away goals modelled in game t.
- ◇ x_t. This denotes the actual home goals in game t.
- ◇ y_t. This denotes the actual away goals in game t.
- ◇ λ_t^H. This denotes the expected home goals in game t.
- ◇ λ_t^A. This denotes the expected away goals in game t.
- ◇ $\alpha_{l,t}$. This denotes the attacking strength of team l for game t.
- ◇ $\beta_{l,t}$. This denotes the defensive strength of team l for game t.
- ◇ $\phi_{l,t} = \frac{1}{\beta_{l,t}}$. This is an alternate parameter for the defensive strength of team l for game t.
- ◇ γ_t. This denotes a home ground advantage parameter, for game t.
- ◇ $\tilde{p}_{l,t}^{\alpha}$ and $\tilde{q}_{l,t}^{\alpha}$ are the hyper parameters for the prior distribution of $\alpha_{l,t}$.
- ◇ $\tilde{p}_{l,t}^{\beta}$ and $\tilde{q}_{l,t}^{\beta}$ are the hyper parameters for the prior distribution of $\beta_{l,t}$.
- ◇ $\tilde{p}_{l,t}^{\gamma}$ and $\tilde{q}_{l,t}^{\gamma}$ are the hyper parameters for the prior distribution of $\gamma_{l,t}$.
- ◇ $p_{l,t}^{\alpha}$ and $q_{l,t}^{\alpha}$ are the hyper parameters for the posterior distribution of $\alpha_{l,t}$.
- ◇ $p_{l,t}^{\beta}$ and $q_{l,t}^{\beta}$ are the hyper parameters for the posterior distribution of $\beta_{l,t}$.
- ◇ $p_{l,t}^{\gamma}$ and $q_{l,t}^{\gamma}$ are the hyper parameters for the posterior distribution of $\gamma_{l,t}$.
- ◇ z_t. This denotes the result of game t. The possible values are: [1,0,0] for a home win, [0,1,0] for a draw and [0,0,1] for an away win.
- ◇ ω_w. This denotes the within season forgetting factor.
- ◇ ω_b. This denotes the between season forgetting factor.
- ◇ ω_h. This denotes the home ground advantage forgetting factor.

To simplify the equations, $\beta_{l,t}$ will be used for the defensive strengths in this section. A lower $\beta_{l,t}$ value is better than a higher one, as this will lower the rate parameter for the opposition goals. Figures 5.1 and 5.2 show the dependence structure in the model after the extend step and after the observations are made respectively, with the inclusion of the hyper parameters of the distributions. The expected goals in each game are modelled by

$$\begin{tikzcd}
& \alpha_{i,t} \sim \mathrm{Gamma}(\tilde{p}^{\alpha}_{i,t}, \tilde{q}^{\alpha}_{i,t}) \arrow[d] & \alpha_{j,t} \sim \mathrm{Gamma}(\tilde{p}^{\alpha}_{j,t}, \tilde{q}^{\alpha}_{j,t}) \arrow[d] \\
\gamma_t \sim \mathrm{Gamma}(\tilde{p}^{\gamma}_{t}, \tilde{q}^{\gamma}_{t}) \arrow[r] & X_t \arrow[r, dash] & Y_t \\
& \beta_{j,t} \sim \mathrm{Gamma}(\tilde{p}^{\beta}_{j,t}, \tilde{q}^{\beta}_{j,t}) \arrow[u] & \beta_{i,t} \sim \mathrm{Gamma}(\tilde{p}^{\beta}_{i,t}, \tilde{q}^{\beta}_{i,t}) \arrow[u]
\end{tikzcd}$$

Figure 5.1: Dependence structure of the Bayesian dynamic football model, showing the prior distributions after the extend step.

$$\begin{tikzcd}
& \alpha_{i,t} \sim \mathrm{Gamma}(p^{\alpha}_{i,t}, q^{\alpha}_{i,t}) \arrow[d] & \alpha_{j,t} \sim \mathrm{Gamma}(p^{\alpha}_{j,t}, q^{\alpha}_{j,t}) \arrow[d] \\
\gamma_t \sim \mathrm{Gamma}(p^{\gamma}_{t}, q^{\gamma}_{t}) \arrow[r] & x_t \arrow[r, dash] & y_t \\
& \beta_{j,t} \sim \mathrm{Gamma}(p^{\beta}_{j,t}, q^{\beta}_{j,t}) \arrow[u] & \beta_{i,t} \sim \mathrm{Gamma}(p^{\beta}_{i,t}, q^{\beta}_{i,t}) \arrow[u]
\end{tikzcd}$$

Figure 5.2: Dependence structure of the Bayesian dynamic football model, showing the posterior distributions after the observations are made.

$$X_t|\lambda_t^H \sim \mathrm{Poisson}(\lambda_t^H) \quad \lambda_t^H = \alpha_{h_t}\beta_{a_t}\gamma_t$$
$$Y_t|\lambda_t^A \sim \mathrm{Poisson}(\lambda_t^A) \quad \lambda_t^A = \alpha_{a_t}\beta_{h_t}$$

The expected goals can be used to attain probabilities of a home win, draw and away win by using the Skellam distribution [38]. This distribution looks at the difference of two independent Poisson random variables. For $t = 2$ to $t = 38$ in each season, the within season forgetting parameter, ω_w, is used to form the prior distributions of the strength parameters. The home ground advatage forgetting parameter, ω_h, is used to form the prior distributions of the home ground advantage parameter. After an observation of a game is made, the hyper parameters for the teams not involved in the game remain unchanged. For the home and away teams involved in this game, t, the updates using ω_w and ω_h are

$$
\begin{aligned}
\pi(\alpha_{i,t}) &\sim \text{Gamma}(\tilde{p}^{\alpha}_{i,t}, \tilde{q}^{\alpha}_{i,t}) = \text{Gamma}(\omega_w p^{\alpha}_{i,t-1}, \omega_w q^{\alpha}_{i,t-1}) \\
\pi(\beta_{i,t}) &\sim \text{Gamma}(\tilde{p}^{\beta}_{i,t}, \tilde{q}^{\beta}_{i,t}) = \text{Gamma}(\omega_w p^{\beta}_{i,t-1}, \omega_w q^{\beta}_{i,t-1}) \\
\pi(\alpha_{j,t}) &\sim \text{Gamma}(\tilde{p}^{\alpha}_{j,t}, \tilde{q}^{\alpha}_{j,t}) = \text{Gamma}(\omega_w p^{\alpha}_{j,t-1}, \omega_w q^{\alpha}_{j,t-1}) \\
\pi(\beta_{j,t}) &\sim \text{Gamma}(\tilde{p}^{\beta}_{j,t}, \tilde{q}^{\beta}_{j,t}) = \text{Gamma}(\omega_w p^{\beta}_{j,t-1}, \omega_w q^{\beta}_{j,t-1}) \\
\pi(\gamma_t) &\sim \text{Gamma}(\tilde{p}^{\gamma}_{t}, \tilde{q}^{\gamma}_{t}) = \text{Gamma}(\omega_h p^{\gamma}_{t-1}, \omega_h q^{\gamma}_{t-1})
\end{aligned}
$$

That is, the hyper parameter values for the prior distributions are equal to the hyper parameter values for the posterior distributions after the previous game, $t-1$, multiplied by ω_w or ω_h. This step keeps the mean of the distribution the same, but increases the variance, as shown in Chapter 4. The formulation for the posterior distributions given the observations is more complicated than shown in equation 4.3, due to there being more than one parameter involved in the likelihoods of the home and away goals for football. Considering the likelihood of the away goals, $\alpha_{j,t}$, the attacking strength of the away side in game t, the posterior update comes from

$$
\begin{aligned}
\pi(\alpha_{j,t}|y_1, y_2, ..., y_t) &\underset{\sim}{\propto} \pi(\alpha_{j,t}) \times L(\alpha_{j,t}\hat{\beta}_{i,t}) \\
&\propto \alpha_{j,t}^{\tilde{p}^{\alpha}_{j,t}-1} e^{-\tilde{q}^{\alpha}_{j,t}\alpha_{j,t}} \times (\alpha_{j,t}\hat{\beta}_{i,t})^{y_t} e^{-\alpha_{j,t}\hat{\beta}_{i,t}} \\
&\propto \alpha_{j,t}^{\tilde{p}^{\alpha}_{j,t}+y_t-1} e^{-\alpha_{j,t}(\tilde{q}^{\alpha}_{j,t}+\hat{\beta}_{i,t})}
\end{aligned}
\tag{5.1}
$$

This gives a $\alpha_{j,t}$ ~Gamma$(\tilde{p}^{\alpha}_{j,t} + y_t, \tilde{q}^{\alpha}_{j,t} + \hat{\beta}_{i,t})$ distribution, where the updates are

$$
\begin{aligned}
p^{\alpha}_{j,t} &\leftarrow \tilde{p}^{\alpha}_{j,t} + y_t = \omega_w p^{\alpha}_{j,t-1} + y_t \\
q^{\alpha}_{j,t} &\leftarrow \tilde{q}^{\alpha}_{j,t} + \hat{\beta}_{i,t} = \omega_w q^{\alpha}_{j,t-1} + \beta_{i,t-1}
\end{aligned}
$$

The likelihood in Equation 5.1 uses an estimate for $\beta_{i,t}$, since this value has not been updated yet. This is $\beta_{i,t-1} = \frac{p^{\beta}_{i,t-1}}{q^{\beta}_{i,t-1}}$. The terms that aren't dependent on $\alpha_{j,t}$ can be treated as constants, and so the $(\hat{\beta}_{i,t})^{y_t}$ term is removed when considering the proportionality. It can be seen that the more goals the away team scores, the larger the value of the attacking strength of the away team, $\alpha_{j,t}$, will be updated to, which is expected. The expected goals are $\hat{\alpha}_{j,t}\hat{\beta}_{i,t}$, which is equal to $\alpha_{j,t-1}\beta_{i,t-1}$. The lower the defensive ability of the home team, meaning the higher the value of $\hat{\beta}_{i,t}$, the higher the expected goals of the away team.

For $\beta_{i,t}$, the defensive strength of the home side in game t, the posterior update comes from

$$
\begin{aligned}
\pi(\beta_{i,t}|y_1, y_2, ..., y_t) &\underset{\sim}{\propto} \pi(\beta_{i,t}) \times L(\hat{\alpha}_{j,t}\beta_{i,t}) \\
&\propto \beta_{i,t}^{\tilde{p}^{\beta}_{i,t}-1} e^{-\tilde{q}^{\beta}_{i,t}\beta_{i,t}} \times (\hat{\alpha}_{j,t}\beta_{i,t})^{y_t} e^{-\hat{\alpha}_{j,t}\beta_{i,t}} \\
&\propto \beta_{i,t}^{\tilde{p}^{\beta}_{i,t}+y_t-1} e^{-\beta_{i,t}(\tilde{q}^{\beta}_{i,t}+\hat{\alpha}_{j,t})}
\end{aligned}
$$

This gives a $\beta_{i,t} \sim$Gamma$(\tilde{p}^{\beta}_{i,t} + y_t, \tilde{q}^{\beta}_{i,t} + \hat{\alpha}_{j,t})$ distribution, where the updates are

$$p^{\beta}_{i,t} \leftarrow \tilde{p}^{\beta}_{i,t} + y_t = \omega_w p^{\beta}_{i,t} + y_t$$
$$q^{\beta}_{i,t} \leftarrow \tilde{q}^{\beta}_{i,t} + \hat{\alpha}_{j,t} = \omega_w q^{\beta}_{i,t} + \alpha_{j,t-1}$$

Here we can see that the more goals the away team concedes, the larger their defensive strength value updates to, meaning the lower their ability in the model. The other parameter updates made after each score is observed are

$$p^{\alpha}_{i,t} \leftarrow \tilde{p}^{\alpha}_{i,t} + x_t = \omega_w p^{\alpha}_{i,t-1} + x_t$$
$$q^{\alpha}_{i,t} \leftarrow \tilde{q}^{\alpha}_{i,t} + \hat{\beta}_{j,t}\hat{\gamma}_t = \omega_w q^{\alpha}_{i,t-1} + \beta_{j,t-1}\gamma_{t-1}$$
$$p^{\beta}_{j,t} \leftarrow \tilde{p}^{\beta}_{j,t} + x_t = \omega_w p^{\beta}_{j,t-1} + x_t$$
$$q^{\beta}_{j,t} \leftarrow \tilde{q}^{\beta}_{j,t} + \hat{\alpha}_{i,t}\hat{\gamma}_t = \omega_w q^{\beta}_{j,t-1} + \alpha_{j,t-1}\gamma_{t-1}$$
$$p^{\gamma}_{t} \leftarrow \tilde{p}^{\gamma}_{t} + x_t = \omega_h p^{\gamma}_{t-1} + x_t$$
$$q^{\gamma}_{t} \leftarrow \tilde{q}^{\gamma}_{t} + \hat{\alpha}_{i,t}\hat{\beta}_{j,t} = \omega_h q^{\gamma}_{t-1} + \alpha_{i,t-1}\beta_{j,t-1}$$

This model can be run over many seasons. For $t = 1$ in each season, except from the first season modelled, the between season forgetting factor, ω_b, is used in place of ω_w. This is applied to the attacking and defensive strengths based on the teams' last values at the final time point of the previous season where possible, where the team was in the division in the previous season as well. This different forgetting factor is used between seasons since there is a longer break between matches, with many changes usually made that effect the teams more than usual. For example, teams usually sign and release numerous players between seasons, and often hire new managers as well.

5.2 Parameter Estimations

The method outlined so far, of using iterative estimates for the unknown parameters, is called mean field approximation [20]. In this case, only a single iteration is used. An alternative to the mean field approximation method is to use MCMC simulations to estimate the parameters [29]. An example is Gibbs sampling [17], which uses dependent sampling. It is similar to the random walk metropolis algorithm [37], except no samples are rejected in the method and it also requires the conditional distributions to be known. If working with two parameters, α and β, where the conditional distributions $\mathrm{P}(\alpha|\beta)$ and $\mathrm{P}(\beta|\alpha)$ are known, the method is as follows

1. Set initial arbitrary values for α^0 and β^0.

2. Choose the number of iterations to perform, n, e.g $n = 10,000$.

3. For i from 1 to n. Sample from $\alpha^i \sim \mathrm{P}(\alpha|\beta^{i-1})$ and record this value. Then sample from $\beta^i \sim \mathrm{P}(\beta|\alpha^i)$.

The samples of α and β can then be used to approximate their distributions. A large enough value of n should be used so that convergence is reached. The process needs to be re-done from scratch to re-estimate parameters after each new observation is made. This method is therefore more convoluted than mean field approximation and much more time consuming. Gibbs Sampling also only works for stationery models and so this methodology would not work for the models proposed here using forgetting factors. MCMC methods for dynamic models are much complex [29]. For these reasons I have only used a mean field approximation method to estimate the unknown parameters, using their expectations from their updates at the previous time.

5.3 Models Created

In this section I compare two different dynamic model approaches. First, I explain the set up a model that uses consistent nuisance parameters throughout all seasons. This includes using a constant within season forgetting factor, ω_w, for all seasons. Following this, I then detail a model that uses a model averaging approach for ω_w. By assessing both models it can be seen whether or not it is beneficial to allow the rate of within season forgetting used within the seasons to differ based on the amount of instability seen in the seasons up to the point of predicting. Constant ω_b and ω_h factors are used throughout for both dynamic models.

The ten seasons of the English Premier League, from the 1996/97 to 2004/05 seasons, are used to decide the nuisance parameters of the models, with the final fifteen seasons then being used to assess their performances, from 2005/06 to 2019/20. These respective data-sets may be referred to as the training and test sets used. The data used to set the model parameters was also truncated. For these models, there is a slight difference to the set up used to explain the theory so far. In order to make the defensive strengths more interpretable and have values that increase as the defensive abilities are believed to increase, the setup of the models use a $\phi_{l,t}$ parameter, which is the reciprocal of $\beta_{l,t}$. The model set ups are then

$$X_t|\lambda_t^H \sim \text{Poisson}(\lambda_t^H) \quad \lambda_t^H = \frac{\alpha_{h_t}}{\phi_{a_t}}\gamma_t$$

$$Y_t|\lambda_t^A \sim \text{Poisson}(\lambda_t^A) \quad \lambda_t^A = \frac{\alpha_{a_t}}{\phi_{h_t}}$$

The nuisance parameters referred to are the ones that need to be set at the start of the model. These are

- The initial $\alpha_{l,0}$ parameters for the teams in the first season.
- The initial $\phi_{l,0}$ parameters for the teams in the first season.

- The initial γ_0 parameters in the first season.
- The initial $\alpha_{l*,0}$ parameters for the promoted teams in each season.
- The initial $\phi_{l*,0}$ parameters for the promoted teams in each season.
- The ω_w value for each season.
- The ω_b value.
- The ω_h value.

The initial values for $\alpha_{l,0}$, $\beta_{l,0}$ and γ_0 for the teams in the first season are not of much concern, since when training a model, this first season, or the beginning of it at least, can be treated as burn in whilst the parameters converge to what would be expected. I therefore set all of these parameter to initially have Gamma$(20, 20)$ distributions. For the other parameters I took two different approaches outlined below.

5.3.1 Model 1 - Constant Within Season Forgetting Factor

The first model was set up using the same fixed nuisance parameters for each season. The values to used were determined by first using the Optim function in R, with a BFGS method [34]. The parameters were optimised against a cumulative log-score for the data from the 1996/97 to 2004/05 seasons. The log-score, seen in equation 3.1, was used since it penalises heavier. The parameters found can be seen in Table 5.1.

From the optimisation, it is interesting to see that a high within season forgetting factor is seen, but the between season forgetting factor is much smaller. This discrepancy supports the need for using a separate forgetting factor between seasons, due to larger changes effecting the teams' strengths. The high within season factor suggests that the seasons are generally stable with not much forgetting required for scores further in the past. The large home ground advantage forgetting factor suggests that this advantage is stable over time.

The values obtained by the optimisation for the hyper parameters of the attacking strengths of the promoted teams seems too large. If the parameter values are too large, they won't be affected enough by the observations when being updated. The hyper parameters for the defensive strength values do seem sensible though. Because of this, I instead used values of

- $p^{\alpha}_{l*,0} = 32$
- $q^{\alpha}_{l*,0} = 39$
- $p^{\beta}_{l*,0} = 39$
- $q^{\beta}_{l*,0} = 32$

and so the hyper parameters for the attacking strengths are the opposites of the ones used for the defensive strengths.

Season	ω_w	ω_b	ω_h	$p^{\alpha}_{l*,0}$	$q^{\alpha}_{l*,0}$	$p^{\beta}_{l*,0}$	$q^{\beta}_{l*,0}$
All Seasons	0.9986	0.5740	0.9983	148.6252	248.2927	39.2067	32.0168

Table 5.1: Table of nuisance parameters optimised using the ten seasons of the Premier League from 1995/96 to 2004/05, for a model with a single within season forgetting factor

5.3.2 Model 2 - Model Averaging for the Within Season Forgetting Factor

The second model that I considered was using a model averaging approach for ω_w. This means that various models are ran using different ω_w values. The final output is obtained by an average of these models, based on the evidence in favour of each. I first considered the parameters to be included for the model and range of ω_w values to use. In order to do this, I again used Optim in R with a BFGS method [34], to optimise the parameters of a single model against the cumulative log score. Separate ω_w values were used for each season. Table 5.2 shows the results obtained from the opimisation. The hyper parameters for the promoted teams were not used for this optimisation, to reduce the parameters to optimise over.

From Table 5.2 it can be seen that the within season forgetting factors are all still very high individually, with a considerably smaller between season factor, though it is not as small as was seen in Table 5.1. This difference in ω_b shows that this parameter is more sensitive to change. This is understandable since it is applied much in the model, and so there are less instances to optimise over. The home ground advantage factor has also not changed much from what was seen in Table 5.1, but it is now seen to be 1 exactly. The smallest within season factors are seen for the 1996/97 and 1998/99 seasons. This shows that these seasons were more unstable, with more unexpected results present. The 1996/97 season stands out, since this was the season Manchester United won on 75 points, the lowest points total of a team that has won the Premier League [31].

Given the results in Table 5.2, a model averaging method was set up to predict with five ω_w values, ranging from 0.96 to 1. A ω_b value of 0.8 was used, along with a ω_h value of 0.999, and the same promoted teams hyper parameters used in Model 1.

An estimation was made for the evidence in the model averaging method for football, which is used to assign the weights to each of the forgetting factors. The predictive distribution described in equation 4.4 is made more complex for the football model, given the extra parameters. All strength parameters and the home ground advantage parameter would need to be integrated out. The estimation used is

$$Z_{1:t} = \sum_{t*=1}^{t} \sum_{k=1}^{3} P_{k,t*}^{z_{k,t*}} (1 - P_{k,t*})^{1-z_{k,t*}}$$

where $P_{k,t*}$ is calculated in the model using the Skellam distribution [38]. This is the probability modelled that the result of game $t*$ was k. This approximation for the evidence is simple, since there are only three possible outcomes for each game, a home win, a draw

Season	ω_w	ω_b	ω_h
1995/96	0.9928	0.8721	1
1996/97	0.9790	0.8721	1
1997/98	0.9996	0.8721	1
1998/99	0.9725	0.8721	1
1999/00	0.9939	0.8721	1
2000/01	0.9961	0.8721	1
2001/02	0.9984	0.8721	1
2002/03	0.9959	0.8721	1
2003/04	0.9944	0.8721	1
2004/05	0.9973	0.8721	1

Table 5.2: Table of nuisance parameters optimised using the ten seasons of the Premier League from 1995/96 to 2004/05, for a model using a model averaging approach for the within season forgetting factor

or an away win.

5.3.3 Method Results

Here I look at the results obtained from running the two dynamic models for the final fourteen seasons, as well as the static Bayesian sequential model. The static model is the same one from Chapter 3, with a couple of key changes. The model uses data from all previous seasons, and includes the same hyper parameters for the promoted teams as both of the dynamic models. The years included in the plots are the years the season ended, for example 2020 represents the 2019/20 season.

Looking at the model averaging approach first, Figure 5.3 shows the evolution of the different weights applied to each forgetting factor for each season. It can be seen how these weights vary a lot over time. Most commonly, a forgetting factor of 1 is seen to have the highest weight, showing that it is best that forgetting is not included in making forecasts at those moments, but this is not always the case. The 2006/07 and 2015/16 are seen to be the most unstable. The 2015/16 stands out, as this was the year where Leicester City famously won the league, after they were big outsiders at the beginning of the campaign. One bookmaker had them priced as high as 5000/1 to become champions [14]. During this season, lots of the top teams that were much more fancied to win the league had inconsistent campaigns, which goes some way to explaining the results of the weightings for this season. Seasons can also be seen that seem stable for most of the season, but then unstable in certain periods. This includes the 2013/14 season, which was unstable at the start and the 2008/09 season which became unstable at the end of the season.

In Figure 5.4, the strengths of all teams for the 2006/07 season are shown using the model averaging model. This shows the instability for this season. The most striking of these are the defensive strengths of Chelsea and Liverpool. Chelsea's defensive strength is seen to decrease drastically around halfway through the season, before quickly increasing again. Liverpool's defensive strength was seen to decrease at the beginning of the season, before suddenly increasing to the best levels in the league, along with Chelsea. This season was well known for West Ham's upturn in form towards the end of the season. They managed to survive, after Carlos Tevez scored a number of goals towards the end of the season. This can be seen in the attacking strengths, whereby it decreased to one of the worst levels in the league, before generally increasing thereafter. A feature of the model averaging model is that it can adapt to unstable seasons such as these by giving more weight to smaller forgetting factors, and so allowing the strengths of the teams to fluctuate more quickly. This is not the case in the single within season forgetting factor model, and other models discussed in Chapter 2 in the literature review, such as the Dixon and Coles model [9].

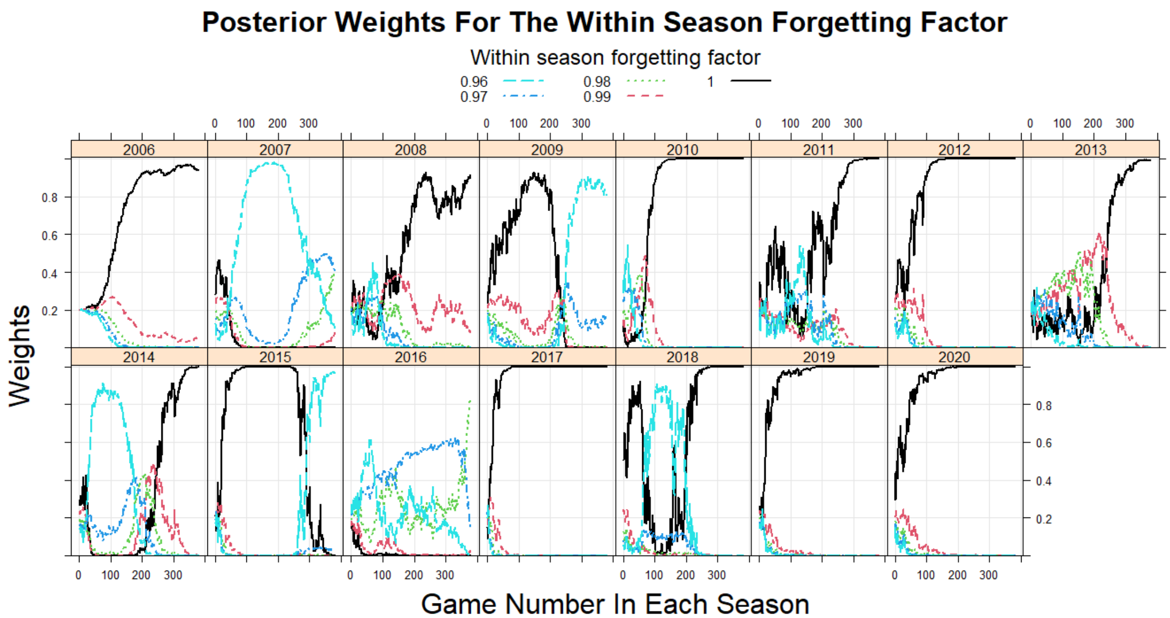

Figure 5.3: Evolution of the weights for each within season forgetting factor value used in the model averaging method, for the 2005/06 to 2019/20 seasons

Figure 5.5 shows the ranked probability scores (see equation 3.2) of the different Bayesian models used for the 2005/06 to 2019/20 seasons. The RPS was used to assess the models, as was done for the static models comparisons in Chapter 3, since it treats match results as ordinal variables. From Figure 5.5, it can be seen that the static model performs the worst, as expected. This is little difference in the first few seasons, but in the later seasons a clear difference can be seen, as the effect the forgetting parameters increases.

The two dynamic models are seen to perform similarly for most seasons. Some seasons the single ω_w model performs the best, such as in 2018/19, whilst in others, such as 2015/16 the model averaging approach is seen to work the best. The model averaging approach can be seen to be significantly better in 2015/16, when considering the usual

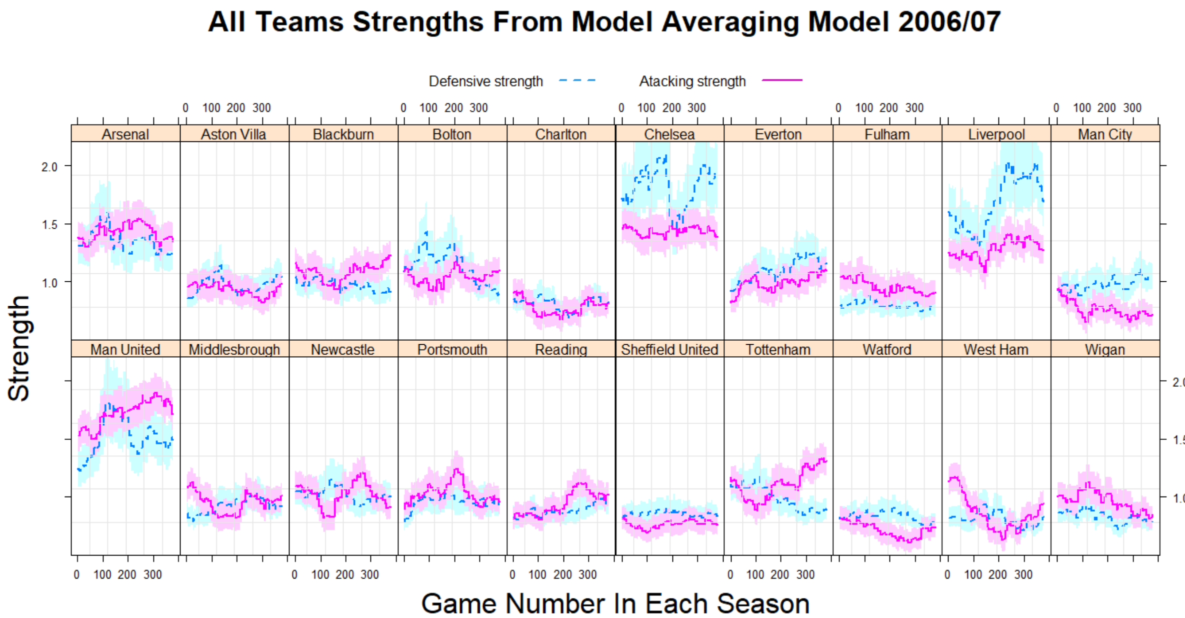

Figure 5.4: The strengths of all teams from the 2006/07 season using a dynamic Bayesian state space model with a model averaging method for the within season forgetting factor

difference between the two models over the other seasons. This seems significant, since Figure 5.3 showed that this season was seen to be quite unstable given the weights applied to the different ω_w values. This suggests that a model averaging approach could work best for unstable seasons, since greater weight can be given to models with lower ω_w values. Looking at the 2006/07 season however, little difference can be seen between the two models, even though Figure 5.3 also points to this being an unstable season also. It would be interesting to compare the model performances over more unstable seasons, perhaps using data from different leagues. If the model averaging model were to always work the best, this would suggest that the stability observed in the past for a given season is indicative of the stability going forwards, but this may not necessarily be the case.

5.4 Strengths Plots

In this section I will focus on the strengths plots created from both of the dynamic models created in Section 5.3. These plots can be used to gain insights into the different teams' performances over the years. This another one of the useful and interesting aspects of the models.

Firstly, a plot of the attacking and defensive strengths using a single within season forgetting factor is considered. This can be seen in Figure 5.6 for Manchester United from 2005/06 to 2019/2020. It can be seen from this plot that although there are periods where the strengths increase or decrease a lot, the change of the strengths from game to game is

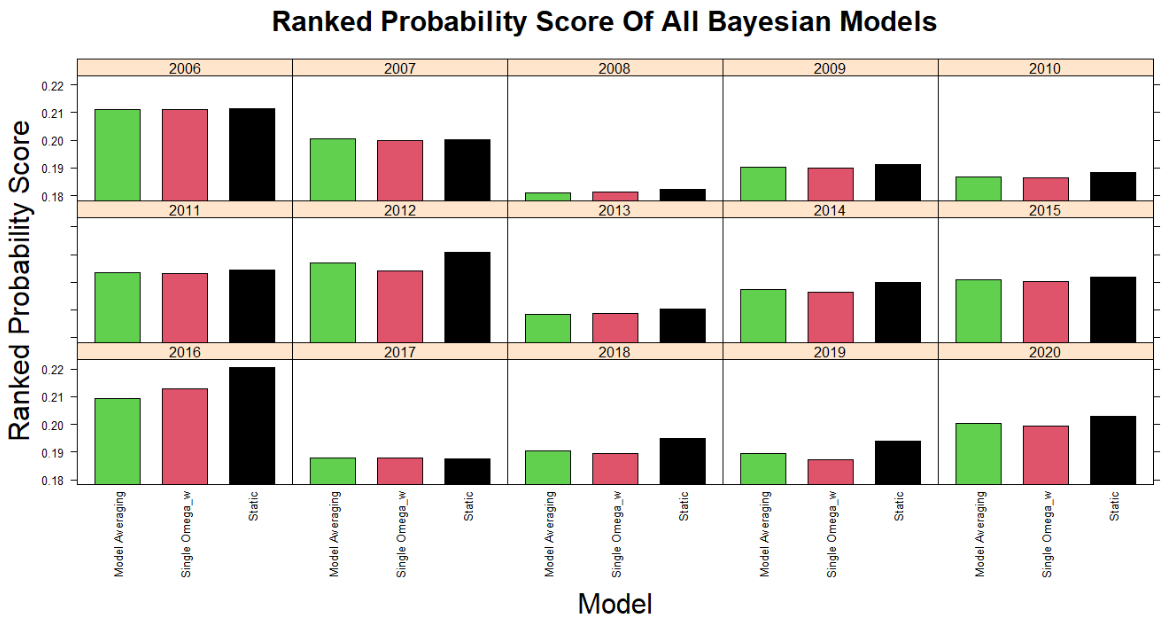

Figure 5.5: Ranked probability scores for the static, single within season forgetting factor and model averaging Bayesian models, for the 2005/06 to 2019/20 seasons

steady, especially as the seasons increase when more data is incorporated into the model. This makes it a good plot to look at the style of the team and the changes in strengths over time. For most seasons, the attacking strength is greater than the defensive strength, showing the attacking nature of Manchester United. For the 2005/06 to 2012/13 seasons, Sir Alex Ferguson was the team's manager, the most successful in their history [32]. During this period, the defensive strength was higher for most of the 2007/08 season until around halfway through the 2009/10 season. It is interesting that in 2008 and 2009 Manchester United reached the UEFA Champions League Final, Europe's most prestigious football competition, during this time of high defensive strength [12]. The very high defensive strength reached in 2008/09 is due to a Premier League record of 14 consecutive games without conceding a goal [13]. Interestingly, this coincides with the big shift in weights for forgetting factors for this season, seen in Figure 5.3. Since Sir Alex Ferguson left the club, the attacking strength was seen to steadily decrease for several seasons. From May 2018 to December 2019, defensive coach Jose Mourinho was the manager [1], and this defensive style can be seen from the strengths in the model. Under his guidance they won the 2017 UEFA Europa League in Stockholm, Europe's secondary international club competition [25]. After a fall in the defensive strength observed in the first half of the 2018/19 season however, Mourinho was sacked and replaced by ex-player Ole Gunnar Solskjaer who has brought an attacking style of play back [23]. The strengths however are still much lower than seen during the Sir Alex Ferguson era, and the club is yet to reach that level of success since in the Premier League.

Next, Figure 5.7 shows the attacking and defensive strengths using a model averag-

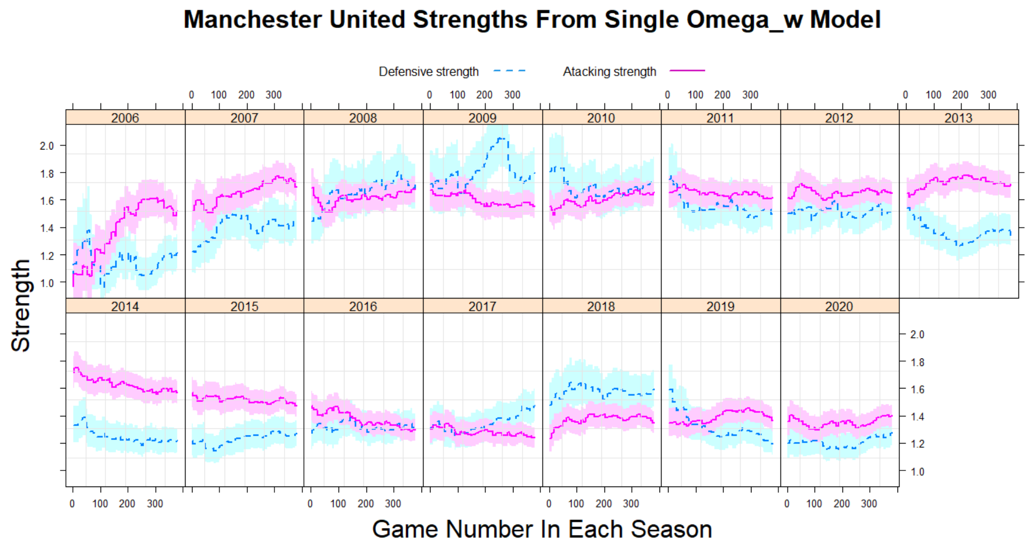

Figure 5.6: The strengths of Manchester United shown from 2005/06 to 2019/2020 using a dynamic Bayesian state space model using a single within season forgetting factor

ing approach for the within season forgetting factor, using Liverpool as the example team. This plot can be immediately seen to be different to Figure 5.6, due to the changes in how much the strengths change from game to game at different times. For instance, the last four seasons in 5.7, from 2016/17 to 2019/20 show little movement from game to game, and the strengths are seen to remain quite steady. This is because the model averaging model gave a very high weight to a ω_w value of 1 for these seasons, meaning no within season forgetting. The strengths have increased steadily over time, as Liverpool have become stronger in the Premier League under Jürgen Klopp, leading to them becoming runners up in 2018/19 and then champions in 2019/20 [26], but haven't increased as fast as they would have done if less within season forgetting was used. In contrast, the strengths for the 2006/07 and 2015/16 seasons can be seen to very much more erratic, where lower values of ω_w were used. The 2006/07 season especially is interesting, as here the defensive strength reached a value of 2, higher than Manchester United's strength for 2008/09 seen in Figure 5.6. Liverpool did have a strong defensive run in this season, conceding just once in nine games during their peak [24], but it wasn't quite the record breaking run that Manchester United had. Nevertheless, given the low within season forgetting used for this season in the model averaging model, the defensive strength updated quite quickly to this high value.

5.5 Summary

In this chapter a dynamic Bayesian state space model for football was explained, expanding on the more generic model discussed in Chapter 3. This model includes attacking

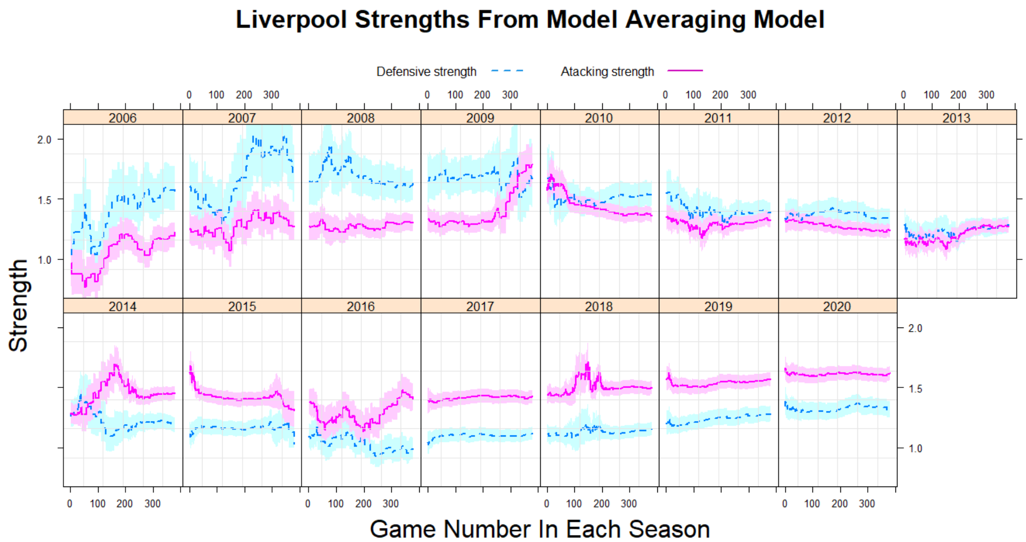

Figure 5.7: The strengths of Liverpool shown from 2005/06 to 2019/2020 using a dynamic Bayesian state space model with a model averaging method for the within season forgetting factor

and defensive strength parameters, as well as a home ground advantage parameter. The updates to be made to the hyper parameters for each was shown, which depend on the conditional distributions. One of the complications involves estimating parameters that are unknown when updating. This model uses a mean field approximation approach by simply using the value at the previous time. This method is believed to be sufficient and is much simpler and less time consuming than applying an MCMC approach instead [29]. Two different dynamic models were then discussed, one which uses a single constant within season forgetting factor, and another which uses a model averaging approach for this variable. Optimisations were performed using Premier League data from 1995/96 to 2004/05 to decide which parameter values to use in the models. The 14 following seasons, from 2005/06 to 2019/20 were then used to compare the models. The weights applied to each forgetting factor in the model averaging approach was first observed, which showed that most seasons had a very high weight given to a factor of 1. The 2006/07 and 2015/16 were the biggest exceptions seen, due to more unexpected results being present in these seasons, for which the seasons may be termed 'unstable'.

Ranked probability scores for each season were analysed for each model, alongside the static Bayesian model which used no forgetting factors. The two dynamic models were seen to perform considerably the best, especially in the later seasons by which time the forgetting parameters have had a big effect. There is little difference in performance seen however between the two dynamic models. The model averaging model does perform better in the 2015/16, a season seen to be unstable from the weights of the model averaging model, by quite a clear distance, which suggests that this model may be best for forecasting results in more unstable seasons. However the 2006/07 was also seen to be unstable, yet both dynamic models predicted similarly for this season. It would be interesting to compare the performances of both models for more unstable seasons, perhaps using data from different leagues. Finally, plots for the attacking and defensive strengths from both dynamic models were discussed and explained. Liverpool and Manchester United were used as example teams for the plots.

Chapter 6

Conclusion

To conclude, this report details and shows the usefulness of a dynamic Bayesian state space model for football, using a conjugacy approach. Reliable data of historic scores of the English Premier League was used.

Various static models were discussed and compared, using the Brier score, log score and ranked probability score metrics. The static Bayesian model was seen to give the best averages for each. It also consistently performed the best in each individual season using the ranked probability score, from 1995/96 to 2019/20.

State space models were explained next, in particular a sequential Bayesian method. This included a model averaging approach, using weights from a range of models using different forgetting factor values to give the final outputs.

A Bayesian state space model for football was then discussed. Estimating unknown parameters is an issue in this model, though this was managed by using a mean field approximation approach, using the parameter values at the previous time. This is more efficient and less complex than using Monte Carlo Markov Chain methods [29]. Two different dynamic models were then discussed, one which uses a single constant within season forgetting factor, and another which uses a model averaging approach instead. Data from 1995/96 to 2004/05 was used to decide the nuisance parameters to use.

The 14 seasons, from 2005/06 to 2019/20 were then used the dynamic models, along with a static one also, which included no forgetting. The weights applied to each forgetting factor in the model averaging approach were mainly seen to be 1, although the 2006/07 and 2015/16 were notable exceptions. This is because more unexpected results occurred in these seasons, with the 2015/16 being the famous season when Leicester City won the title, who were big underdogs at the start of the campaign. The ranked probability score was used to assess the performances of the models. This showed that the dynamic models consistently performed much better, especially in the later seasons, by which time the forgetting factors have had a bigger effect. There was however, little difference between dynamic models themselves. The model averaging model did perform better in the 2015/16 by quite a clear distance, which suggests that this model may be best for forecasting results in more unstable seasons. The 2006/07 was also seen to be unstable, yet both dynamic models predicted similarly for this season. The performances of both models for more unstable seasons, perhaps using data from different leagues, could therefore be carried out for further investigation.

Lastly, the strength plots for Liverpool and Manchester United were discussed, using both dynamic models. These gave some insight into the models and showed the usefulness of the strength parameters, which can be used to assess team performances.

Appendix

Appendix A Bayesian Dynamic Modelling of Football Results: Project Specification

Matthew Williams

Project at Lancaster University, supervised by Gareth Ridall

7th June 2020

Motivation

Predicting the outcomes of football games is currently a very popular topic for data scientists. This is because football is the most watched sport in the world, and with that, it is also the most popular sport for betting.

There are several ways of trying to predict results of matches ('Home win', 'Draw', 'Away win'). One of these ways is to use a static model, with parameters that are fixed in the model and do not change over time. This is the classical approach. These kinds of static models have been seen to predict football matches poorly, due to factors, such as the form of the teams, changing over time in reality. These changes can occur over a short period of time.

In this project, I will look to create a dynamic model to predict football results, using a Bayesian approach to update the parameters. This will allow the parameters to change over time, more in line with reality. This should lead to better predictions.

Generating football predictions that can be used to "beat the bookies", by identifying games to bet on with bookmakers to eventually return a profit, is a very difficult task. This is because bookmakers make use of lots of data and have many years of experience. In this project, I will only be using the scores and dates of football games. I will therefore be looking to create predictions that are close to bookmakers' odds, but do not expect to build a method that can "beat the bookies".

Data

The data that will be used for this project is from a website called Football-Data. I will only be using the scores of the matches, along with the dates they occurred to create models and systems. Data on the bookmakers' odds for the games will also be used for comparisons with predictions made.

The data from Football-Data is known to be highly accurate, and any mistakes are corrected quickly. These are some of the advantages of such data regarding football. This is because the popularity of the sport, and the number of keen followers, means that people often submit correct data and if not, errors are quickly identified by others.

The two programming languages I am most familiar with are Python and R. Analysis on the datasets in this project will be done using R. This is because the mathematical and statistical nature of this project lends itself more to R, rather than Python.

Aims

There are many topics to explore for this research area of predicting football results. These are in terms of the analysis and creation of models, as well as comparisons that can be made between different aspects of the game. Given the 12-week timeframe of this project, I have identified three key aims, along with objectives for each that I look to accomplish.

1) To create Bayesian a dynamic model and assess its suitability.

Firstly, I will create a Bayesian dynamic model for predicting football results. I will then assess the suitability of this model. I expect that this will be the aim that will feature most in the end paper and take the most time to work on.

The model I create will use parameters for the home and away teams' strengths in each game, as well as one for the home ground advantage. These parameters will be able to change over time, unlike a classical static model whereby they remain fixed. I will not be using Markov Chain Monte Carlo (MCMC) methods in the dynamic model, as this method will be very time consuming. I will instead look to exploit conjugacy between the prior and posterior distributions in a Bayesian approach. This will mean that the parameters can be updated exactly, without a need for estimation. To calculate the parameter values at a point in time, I will look to use exponential decay when incorporating past information. This will be one of the complexities of creating the dynamic model, as I will have to try and work out how to best apply this feature.

To assess the suitability of the model I create; I will assess which scoring methods should be used. These include the Brier score, ranked probability score (RPS) and log-score. I will judge which method I deem to be most appropriate for assessing football predictions. I will then include a comparison against a classical static model, as this will be a benchmark for the dynamic model to better. To do this, I will create a classical dynamic model, which will be the first task in this project.

Objectives:

- Create a Bayesian dynamic model for predicting football results.
- Identify suitable scoring methods to assess the performances of predictive models for football results.
- Assess the performance of the Bayesian dynamic model, including a comparison against a classical static model.
- Compare the predictive power of the Bayesian dynamic model against bookmakers' odds.

2) Create a system for ranking football teams' performances, like the ELO system used in Chess, and assess its suitability.

The next aim of this project will be to create a rating system for football teams, like the ELO system used in Chess. Ratings systems like ELO have become popular in other sports, such as Baseball, but have not yet been widely applied to football. I will assess whether the system I create is suitable. I will also look to see if accurate predictions of football results are able to be made using the ratings, with comparisons to bookmakers' odds. For use in the predictions of results, less parameters will be included than in the Bayesian dynamic model. It will be interesting to see how much this affects the predictive power.

Objectives:

- To create a rating for football teams' performances, like the ELO system used in Chess.
- To assess the suitability of the rating system created.
- Investigate whether the ratings can be used to create accurate predictions of football results, including comparisons with bookmakers' odds.

3) Compare the performances of teams across the top two divisions in England, Germany, and Spain, as well as the suitability of the Bayesian dynamic model and rating system created.

My final aim is to use the Bayesian dynamic model and rating system created to make comparisons to different situations in football. The angle that I will be take is a comparison between three countries: England, Germany, and Spain. I will be looking at the top two divisions in each of these countries.

England, Germany, and Spain are currently seen to be the best countries for football in Europe in terms of the quality of the teams. These leagues should therefore be well suited to compare against each other. The popularity of football in these countries also means that data is readily available and can be trusted to be up to date and accurate.

The comparisons that I look to make between the divisions will include volatility, suitability of the model and ranking system that I created, and predictions created from them compared to bookmakers' odds. For the volatility, I will be looking at how the change in strengths of the teams varies over time, both within and between seasons.

Objectives:

- To assess the volatility of the change in strengths of teams across time in each division.
- To assess the suitability of a Bayesian dynamic model and rating system for the teams in each division.
- To compare how close predictions made for each division are to bookmakers' odds.

Timeline

June

- Research and implementation of a classical static model for football result predictions.
- Research of Bayesian dynamic models for football result predictions.

July

- Implementation and analysis of a Bayesian dynamic model for football result predictions.
- Research of rating systems for sports, such as the ELO system used in Chess.

August

- Implementation and analysis of a rating system for football teams.
- Comparison of the top two divisions in England, Germany, and Spain.

Changes may be made to the aims and objectives of the project due to the timescale and findings during the research.

Deliverables

- A Bayesian dynamic model for predicting football results in the top two divisions of England, Germany, and Spain.
- The creation of a rating system for football teams in the top two divisions of England, Germany, and Spain.
- An academic paper including an assessment of the model and rating system created, with a comparison between the different divisions analysed.

Bibliography

[1] BBC. Jose mourinho: Man utd confirm former chelsea boss as new manager, 2016. Available at `https://www.bbc.co.uk/sport/football/36363365`.

[2] J.-M. Bernardo. Expected information as expected utility. *Ann. Statist.*, 7(3):686–690, 1979.

[3] G. Boshnakov, T. Kharrat, and I. G. McHale. A bivariate weibull count model for forecasting association football scores. *International journal of forecasting*, 33(2):458–466, 2017.

[4] R. A. Bradley and M. E. Terry. Rank analysis of incomplete block designs: I. the method of paired comparisons. *Biometrika*, 39(3/4):324–345, 1952.

[5] G. W. Brier. Verificartion of Forecasts Expressed In Terms of Probability. *Monthly Weather Review*, 78(1):1–3, 01 1950.

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

[7] R. R. Davidson. On extending the bradley-terry model to accommodate ties in paired comparison experiments. *Journal of the American Statistical Association*, 65(329):317–328, 1970.

[8] A. C. Davison. *Statistical Models*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, 2003.

[9] M. J. Dixon and S. G. Coles. Modelling association football scores and inefficiencies in the football betting market. *Journal of the Royal Statistical Society. Series C, Applied statistics*, 46(2):265–280, 2002.

[10] E. S. Epstein. A scoring system for probability forecasts of ranked categories. *Journal of applied meteorology (1962)*, 8(6):985–987, 1969.

[11] football-data.co.uk. Data files: England, 2020. Available at `http://www.football-data.co.uk/englandm.php`.

[12] footballhistory.org. Champions league, 2020. Available at `https://www.footballhistory.org/tournament/champions-league.html`.

[13] FourFourTwo. Van der sar concedes following 14 clean sheets, 2009. Available at `https://www.fourfourtwo.com/news/van-der-sar-concedes-following-14-clean-sheets`.

[14] Frodsham, Isobel. Football fan who won £20,000 when he drunkenly bet on leicester city to win the premier league says he might go professional after another £760 windfall from a £1 football stake, 2016. Available at https://www.dailymail.co.uk/news/article-4064846/Leicester-City-fan-won-20-000-5000-1-bet-jokes-professional.html.

[15] B. G, K. T, and M. I.G. Are goals poisson distributed? Technical report, Centre for Sports Business, Salford Business School, University of Salford, 2016.

[16] D. Gamerman, T. R. dos Santos, and G. C. Franco. A non-gaussian family of state-space models with exact marginal likelihood. *Journal of time series analysis*, 34(6):625–645, 2013.

[17] S. Geman and D. Geman. Stochastic relaxation, gibbs distributions, and the bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-6(6):721–741, 1984.

[18] M. E. Glickman. Parameter estimation in large dynamic paired comparison experiments. *Journal of the Royal Statistical Society Series C*, 48(3):377–394, 1999.

[19] Global Futbol Training. When are most goals scored in soccer?, 2020. Available at https://www.gftskills.com/when-are-most-goals-scored-in-a-game/.

[20] M. I. Jordan, Z. Ghahramani, T. S. Jaakkola, and L. K. Saul. An introduction to variational methods for graphical models. *Machine Learning*, 37(2):183–233, 1999.

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

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

[23] Manchester United. Ole gunnar solskjaer appointed caretaker manager, 2018. Available at https://www.manutd.com/en/news/detail/ole-gunnar-solskjaer-appointed-caretaker-manager.

[24] Mannion, Dean. Season in review: 2006/07, 2017. Available at https://www.friendsofliverpool.com/2017/07/season-review-200607/.

[25] McNulty, Phil. Ajax 0-2 manchester united, 2017. Available at https://www.bbc.co.uk/sport/football/40004559.

[26] McNulty, Phil. Liverpool win premier league: Jurgen klopp's transformative role, 2020. Available at https://www.bbc.co.uk/sport/football/51751082.

[27] Moncav, Melony. What is elo? an explanation of competitive gaming's hidden rating system, 2020. Available at `https://dotesports.com/general/news/elo-ratings-explained-20565`.

[28] A. Owen. Dynamic bayesian forecasting models of football match outcomes with estimation of the evolution variance parameter. *IMA journal of management mathematics*, 22(2):99–113, 2011.

[29] T. Pang. *An introduction to quantum Monte Carlo methods*. IOP (Series). Release 3. IOP Publishing, Bristol [England] (Temple Circus, Temple Way, Bristol BS1 6HG, UK), 2016.

[30] R. Pollard. Home advantage in football: A current review of an unsolved puzzle. *The Open Sports Sciences Journal*, 1, 06 2008.

[31] Premier League. 1996/97 season review: Man utd make it four, 1997. Available at `https://www.premierleague.com/news/59008`.

[32] Premier League. Manager profile alex ferguson, 2020. Available at `https://www.premierleague.com/managers/344/Alex-Ferguson/overview`.

[33] P. V. Rao and L. L. Kupper. Ties in paired-comparison experiments: A generalization of the bradley-terry model. *Journal of the American Statistical Association*, 62(317):194–204, 1967.

[34] RDocumentation. optim, 2020. Available at `https://www.rdocumentation.org/packages/stats/versions/3.6.2/topics/optim`.

[35] RDRR. polr: Ordered logistic or probit regression, 2020. Available at `https://rdrr.io/cran/MASS/man/polr.html`.

[36] P. Ridall, M. George, and A. Pettit. Fast sequential bayesian analysis of football scores illustrating the evolution of the styles and strengths of each of the british premier league football sides over two decades. Technical report, Lancaster University, UK and Queensland University of Technology, Australia, 2019.

[37] M. N. Rosenbluth. Genesis of the monte carlo algorithm for statistical mechanics. *AIP Conference Proceedings*, 690(1):22–30, 2003.

[38] J. G. Skellam. The frequency distribution of the difference between two poisson variates belonging to different populations. *Journal of the Royal Statistical Society*, 109(3):296–296, 1946.

[39] Sourav. Top 10 most popular sports in the world [updated 2020], 2020. Available at `https://sportsshow.net/top-10-most-popular-sports-in-the-world/`.

[40] SportsPro. Premier league 2019/20 commercial guide: Every club, every sponsor, all the major tv deals, 2020. Available at https://www.sportspromedia.com/analysis/premier-league-2019-20-guide-club-sponsor-tv-rights-deals.

[41] Starlizard. Starlizard. the world's most accurate football predictions, 2020. Available at https://www.starlizard.com/.