GHENT UNIVERSITY

FACULTY OF SCIENCES

Modelling Football Outcomes: A Maximum Likelihood Approach

Daan Nyckees

Master dissertation submitted to
obtain the degree of
Master of Statistical Data Analysis

Promoter: Prof. Dr. Christophe Ley
Department of Applied Mathematics, Computer Science and Statistics

Academic year 2018 - 2019

Modelling Football Outcomes: A Maximum Likelihood Approach

Daan Nyckees

Master dissertation submitted to
obtain the degree of
Master of Statistical Data Analysis

Promoter: Prof. Dr. Christophe Ley
Department of Applied Mathematics, Computer Science and Statistics

Academic year 2018 - 2019

Foreword

First, I would like to thank my promotor, professor Christophe Ley from the department of applied mathematics, computer sciences and statistics for guiding me through the process of completing this master thesis. He gave me enough freedom to structure the project, but was always accessible to steer me into the right direction. I have learnt a lot concerning statistical modelling in sports and I am forever grateful for this knowledge and the opportunity to work on such an interesting topic.

Additionally, I would like to thank the people who believed in me during this journey, especially my girlfriend, parents and friends. Supporting me when necessary and motivating me not only through the process of writing this thesis, but throughout my whole academic career. Furthermore, a special thanks goes to mr. Peter Cosyn for reviewing my dissertation and giving some important tips.

Table of Contents

Abstract

Given the vast popularity of soccer and its increasing economic value, modelling outcomes of soccer matches has an increasing value for spectators, online gamblers, board members of clubs, etc. A great deal of statistical methods have been discussed in the past, and most papers address improvements and modifications over the basic independent Poisson model as suggested by Maher et al. (1982). This thesis compares the different models based on maximum likelihood estimation.

Adding to the Poisson and the discrete Weibull distribution which are predominantly used in the literature, we examine the Conway-Maxwell-Poisson and the Zero-Inflated Poisson distribution for their goodness of fit to the data. Additionally, we examine the use of bivariate distributions with a correlation factor by using a copula. Finally, six modifications were tested, giving a total number of 20 models.

The Ranked Probability Score was applied to assess the differences in predictive power between the models. No significant difference could be found between any of the models.

Chapter 1: Research objectives and outline

1.1 Introduction

Association football, or soccer, is a team sport originated in England, which underwent an immense growth in the twentieth century to become the world's most popular team sport (Sport matters, 1999). Data from the FIFA (Fédération Internationale de Football Association) suggests that 270 million people were involved in football in 2006, of which 265 million were football players (FIFA Big Count, 2007). Furthermore, the cup final of the latest World Cup in Russia was followed live by 1,12 billion people.
The popularity and public interest in the game go hand in hand with the economic aspect.
This follows from a recent surge in cash flow in the last decennium especially in the Premier League, mainly due to foreign investors, such as Russian oil tycoon Roman Abramovich (Chelsea FC) and the late Thai Multi Billionaire Vichai Srivaddhanaprabha (Leicester City FC). This translated in 2 122 million euros of transfer fees in 2018 alone for the Premier League (CIES, Centre International d'Etude du Sport). Additionally, four English clubs made the top 10 of transfer fee spenders worldwide from 2010 to 2018 (CIES, Centre International d'Etude du Sport). This excessive spending ensured a supply of talent from all over the world towards the Premier League, making it one of the most exciting, attractive and thrilling football competitions in the world.

A lot depends on the outcome of the different matches and the position of the teams in the table at the end of the season. However, a game of soccer itself depends heavily on chance (Reep and Hill, 1968). The literature states that football experts are to some extent able to predict the final league table positions despite the subjection of chance (Hill, 1974). Although chance dominates the game (Reep and Hill, 1968), skill rather than chance dominates the performance over a season, giving a rationale for statistical modelling and the exertion in trying to forecast the outcomes of football matches. While statistical modelling techniques have been widely applied in other sports, especially American originated sports such as American football, baseball and basketball, research in soccer analytics is relatively sparse. However, the use of data and modelling starts to take off in football. An example of such a model was initially proposed by Maher et al. (1982), where goals are used directly to model team specific strengths by fitting a Poisson distribution to the data. Inherently, the Poisson distribution fits the data rather well both theoretically and practically. Each team has a large number of attacks with a small probability of actually scoring within the same time interval of a match. Therefore, each team's attacking and defensive performance properties can be modelled by means of the Poisson distribution with a different mean of scoring goals for each team separately. The Skellam distribution (Skellam, 1946) refers to the discrete probability distribution of the difference of two Poisson distributed variables and can subsequently be used to model the score line. While this model gave good preliminary results, a few considerations and modifications were taken into account in subsequent research. Most alterations to the model by Maher et al. (1982) include the adaptation of the independence assumption, the dynamic nature of the team performances and the underlying distribution.

First of all, a bivariate Poisson with a small correlation factor between the teams to account for the speed of the game and the influence of the opposing team was suggested (Maher et al., 1982). Dixon et al. (1997) proposed a modification, later on widely adopted as the Dixon-Coles adjustment, where a dependence parameter is added to account for the low scoring games, while the marginal distributions remain Poisson (Dixon et al., 1997). Additionally, they proposed the construction of a pseudo-likelihood where recent performances have larger weights over performances in the past to adjust for the dynamic nature of team performances. Therein team performances are down weighted exponentially by an additional parameter. Karlis et al. (2003) adopted the bivariate Poisson and added a diagonal inflation component to account for the underestimation of draws (Karlis et al., 2003). Finally, Rue et al. (1998) suggested a Bayesian dynamic generalized linear model to estimate the time dependent skills of all teams in a league (Rue et al., 1998). Which also adds an extra parameter to adjust for the hypothesized overestimation of stronger teams when playing weaker teams.

Dixon et al. (1997) were the first to hypothesize the dynamic nature of team performances, addressing the static nature of the basic double Poisson model by exponentially down weighting the likelihood by a time decay function. In contrast, Koopman et al. (2015) used a non-Gaussian state space model to allow the team strengths to vary stochastically over time. They emphasize that while the attacking and defensive strengths of a team do vary over time, it is a slow process. Therefore, most changes are in the long run such that week to week changes are small and season to season changes are more likely.

While most papers built further on the model with Poisson densities as marginal distributions, Boshnakov et al. (2016) introduced the discrete Weibull probability distribution generated by a renewal count process as an alternative to the aforementioned Poisson distribution. The Weibull count model nests other discrete models such as the Poisson and the Negative Binomial as special cases and can handle both over- and under-dispersed data, giving the model more flexibility. Moreover, inter-arrival times are Weibull distributed instead of exponentially. Furthermore, they model the dependency of the scoring ability of both teams by employing a copula to generate a bivariate distribution (Boshnakov et al., 2016).

Many methods exist to infer the inherent qualities of teams in order to model the outcome of matches. This thesis examined the different methods of forecasting football outcomes with the maximum likelihood approach. Additionally, the Conway-Maxwell-Poisson and the Zero-Inflated Poisson distribution were taken into account as a different underlying distribution to model the goal scoring probabilities next to the popular Poisson and discrete Weibull distributions. Other approaches such as Bayesian statistics, state space modelling and machine learning methods were not taken into account in this comparison.

1.2 Problem Statement and research objective

A lot of work has been done to improve the basic independent Poisson model as suggested by Maher et al. (1982). Nevertheless, there is no benchmark for what method is best for predictive purposes of match outcomes. Either the proposed models were compared to the independent Poisson model, or predictive power was calculated based on betting odds and gambling return. This dissertation compares different models as proposed by the literature based on predictive power, which was calculated with the use of the Ranked Probability Score. Additionally, two extra probability distribution and a new adjustment were investigated.

1.3 Outline: The roadmap through this dissertation

Chapter 2 describes the data and the overall methodology with an in-depth explanation of the different probability distributions used to model the goal scoring probabilities and the adjustments as proposed by the literature.

Results and discussion are presented in chapter 3 and 4 respectively.

Finally, some examples of practical use of the thesis are included in chapter 5.

Chapter 2: Methodology

What follows is a detailed description of the dataset used for the analysis, the methods of estimation of the different parameters and the different models. A total number of 20 models were built for comparison. Differences between the models depend on the probability distribution, the use of a correlation factor and the different adjustments made.

2.1 Data

There has been a recent explosion in data gathering techniques and companies that offer tailored sports data resulting in a massive growth of available data for both player and match analysis. However, the focus in this thesis is subject to data that are freely available and easily accessible. The dataset comprises the outcome for every league match of the Premier League for a period of 10 seasons ranging from 2008 until 2018 with corresponding dates and potential home advantage for every game. The data are freely accessible on
http://www.football-data.co.uk/.
This results in a dataset of 3 800 fixtures with 36 different teams with a total of 10 322 goals scored. Additionally, the odds for the three possible outcomes of seven different bookmakers are available. Later on these will be used to see if a favorable betting return is possible.

Not all 36 teams are present every season in the Premier League. The Premier League and the lower tier Championship have a promotion and relegation system, where the three lowest ranked teams from the Premier League are relegated and the top two teams of the Championship get directly promoted. The 3rd to the 6th ranked team battle for the last remaining promotion ticket in a play-off system. Due to the difference in number of games played, a comparison of goals was made per match. Figure 1 displays this relationship of the goals scored per team per match during the 10 season span. It reflects the difference in teams in terms of goal scoring capabilities where the “big six”, which is a nickname for the top six teams throughout this decade, top the standings.

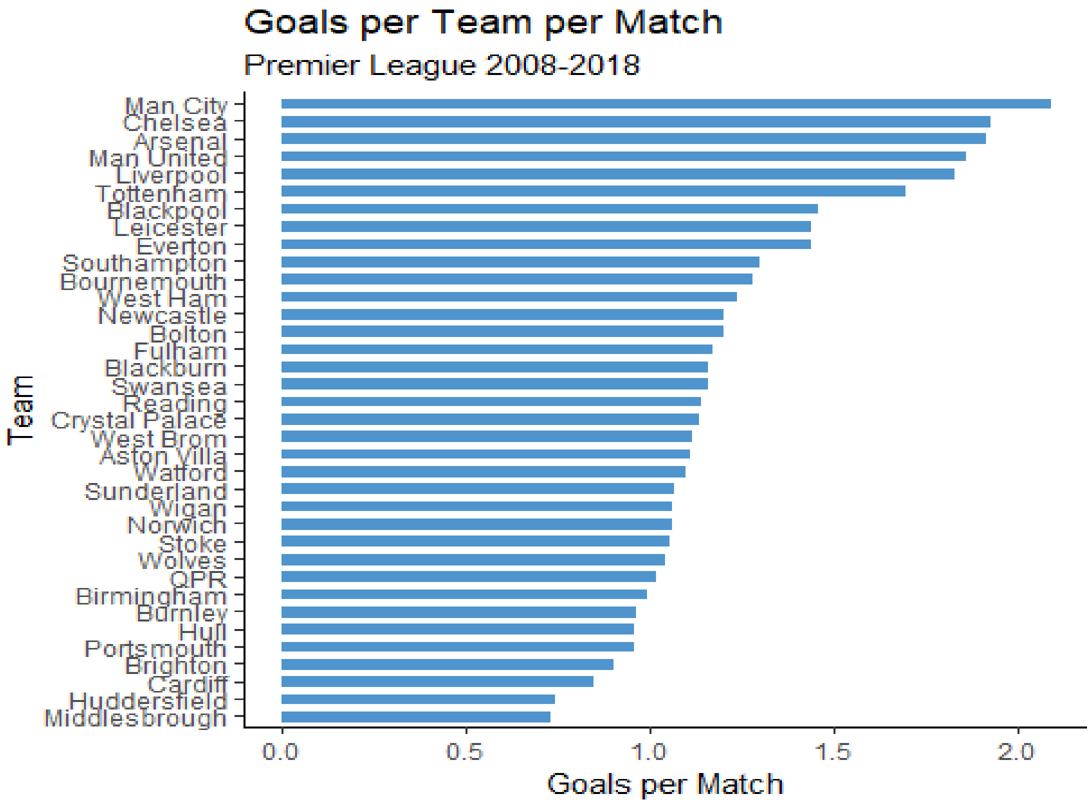

Figure 1. Goals per Team per Match

The dataset makes a distinction in goals scored by the home team or the visiting team. The difference in number of goals per season between the home and away teams is visualized in figure 2. A certain trend is visible: more goals are scored by the home team in comparison to the away team, portraying the possibility of an advantage for the home team to score goals. This topic has already been researched in the past (Pollard et al., 1986). Possible causes for the home advantage include local crowd support, a psychological placebo effect, the elimination of travel fatigue, the familiarity with local conditions, referee bias and the fact that visiting teams may play more defensively (Pollard et al., 1986).

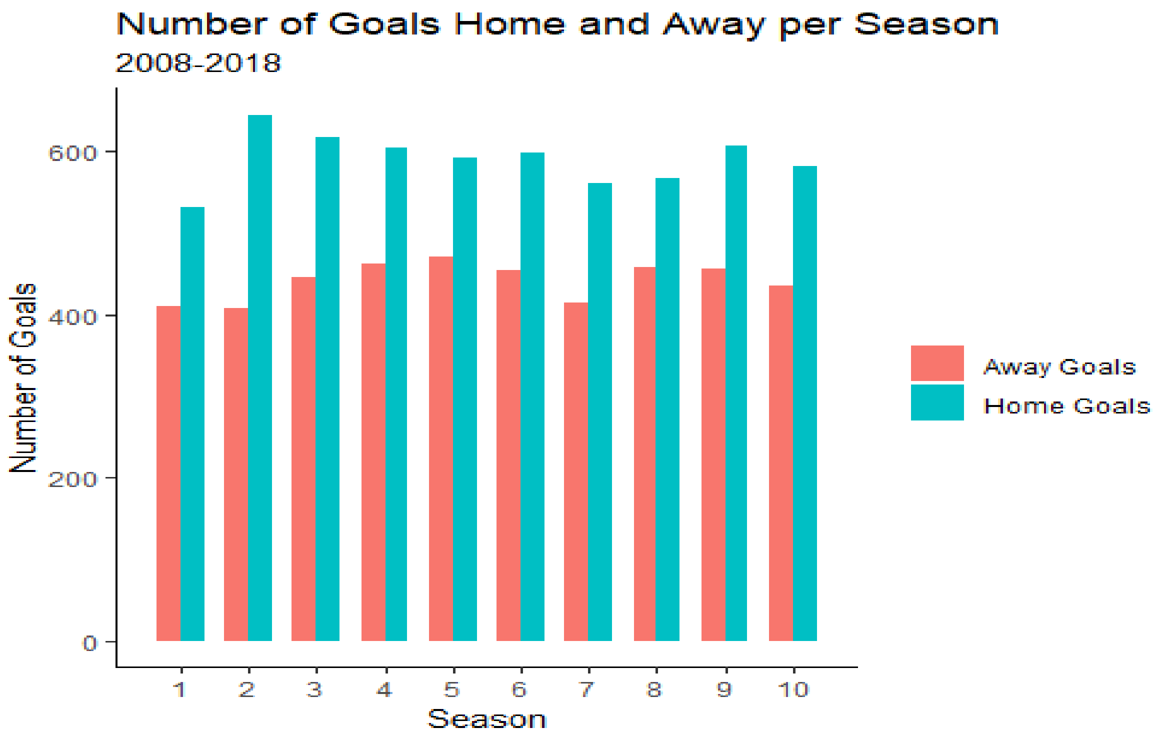

Figure 2. Number of goals home and away per season

To see whether this effect is substantially significant, a Pearson's Chi-square test (1) was conducted and the difference was evaluated based on the Gini Index of dissimilarity (2), which is preferred when dealing with large samples. The null hypothesis of no association between the number of goals scored and the team playing (home or away) applies to both tests.

The Chi-Square test follows asymptotically a Chi-Square distribution. The Gini index results in a value between 0 and 1 and should be lower than 0.03 for a good model fit. Both test statistics are displayed below:

$$(1)\quad \chi^2 = \sum_{j=1}^{J} \frac{\left(n_j - \mu_j\right)^2}{\mu_{j0}}$$

$$(2)\quad D = \sum_{j=1}^{J} \left(\frac{|n_j - \mu_j|}{2n}\right)$$

With n_j being the observed counts and μ_j the expected counts.

The Chi-Square test results in a p-value of 2.2e-16 and D = 0.07198217. Both tests thus yield a significant result.

The distribution of the goals scored both by the home team and the away team are displayed in figure 3. At first glance, a Poisson distribution seems to be a good fit for the data graphically, especially for the goals at home. The goals away seem to require a mixture of Poisson distributions like a Zero-Inflated Poisson model to account for the high number of games which amounted in a 0 score or in 1 goal. But overall, a Poisson distribution as proposed by Maher et al. (1982) seems valid at first sight.

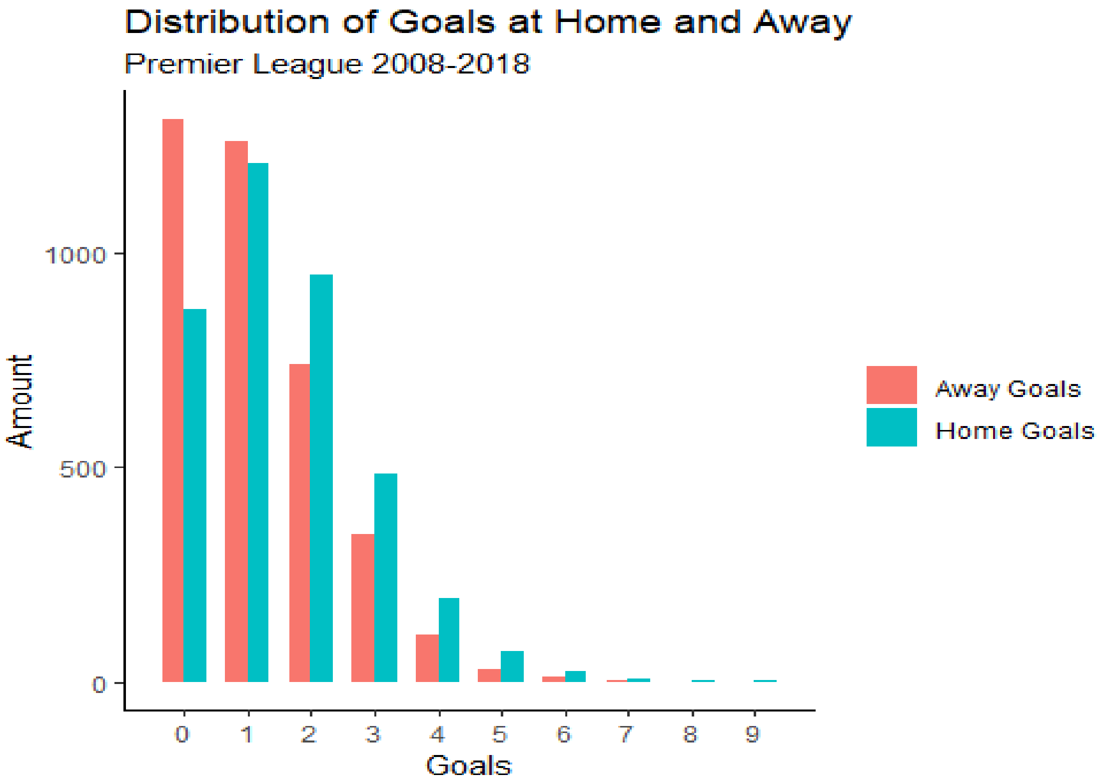

Figure 3. Distribution of goals at home and away

2.2 Methodology

Match outcomes are modelled by means of goal scoring competencies of both teams. Herein are the number of goals scored by the home team a result of the ability of the home team to score goals and the ability of the visiting team to prevent those goals and vice versa for the away team. These attacking and defensive parameters were estimated by means of maximum likelihood and vary across different teams, whereas the home team has an extra parameter to account for the home advantage.

This results in the following general equation:

$$\log_e(\lambda_h) = \alpha_i + \beta_j + \eta$$
$$\log_e(\lambda_a) = \alpha_j + \beta_i$$

The mean number of goals scored in a match by the home and away team are represented by λ_h and λ_a. The parameters α_i and β_j represent the attacking skill and defensive skill of team i and j. The α and β parameters vary across teams and the η parameter depicts the home advantage which is the same across teams. The log link was used to make an additive formula of the parameters such as described in generalized linear model (GLM) procedures.
This generic model as proposed by Maher et al. (1982) gets modified depending on the probabilistic distribution, the addition of a copula and the different adjustments.

Attacking, defending and home advantage parameters are estimated using maximum likelihood, giving the following probability distribution and likelihood function for the basic double Poisson model as an example:

$$P(X_i = x,\ Y_i = y) = \frac{\lambda_h^{\,x} exp(-\lambda_h)}{x!}\,\frac{\lambda_a^{\,y} exp(-\lambda_a)}{y!}$$
$$L(\lambda_h, \lambda_a \mid x_i, y_i) = \prod_{i=1}^{n}\left(\frac{\lambda_h^{\,x_i} exp(-\lambda_h)}{x_i!}\,\frac{\lambda_a^{\,y_i} exp(-\lambda_a)}{y_i!}\right),$$

which results in the following log-likelihood:

$$l(\lambda_h, \lambda_a \mid x_i, y_i) = \sum\left(-\lambda_h - \lambda_a + x_i \log_e(\lambda_h) + y_i \log_e(\lambda_a) - \log_e(x_i!) - \log_e(y_i!)\right)$$

λ_h and $\lambda_a$denote the mean scoring probabilities of the home and away team respectively and the x_i and y_i represent the observed home and away goals in a certain match i.

A sum to zero constraint was added to avoid overparameterization of the model.

$$n^{-1}\sum \alpha_i = 0$$

The data of five seasons and 10 fixtures in the 6th season were used to estimate the parameters for the remaining 6th season. After each fixture, an update of the parameters was performed with the addition of the previous fixture. Thus the training set for the next fixture in a season increases when nearing the end of that season. This analysis was done for 5 sequential seasons, resulting in 1 400 games which could be used for subsequent analysis. Additionally, the Ranked Probability Score was estimated for every match to compare predictive power between the models.
In sample performance was evaluated by comparison of the log-likelihood for the different probability distributions and the Akaike Information Criterion (AIC). Out sample performance was analyzed by means of the Ranked Probability Score.

2.3 Probability distributions

Four probability distributions were compared for model fit for the distribution of goals. First, the basic Poisson distribution was examined. Additionally, a mixture of Poisson distributions, a two-parameter extension of the Poisson and a discrete version of the Weibull distribution were taken into account.

1) *Poisson Distribution*

The Poisson model has been used in various sectors to model the number of occurrences in a given time interval and was firstly introduced by Maher et al. (1982) to model the scoring probabilities of each team separately. It is characterized by a discrete probability distribution that expresses the probability of a given number of events occurring in a fixed interval of time or space. These events happen to occur at a known constant rate and independent of the time since the last event (Haight, 1967), and has the following probability distribution:

$$P(X = x) = \frac{\lambda^x}{x!}e^{-\lambda}$$

Furthermore, the Poisson has the mathematical property of equidispersion, which states that the mean is equal to the variance. Which is in se an attractive mathematical property, but often infeasible in practice.

2) *Conway-Maxwell-Poisson Distribution*

Even though the Poisson distribution has some attractive properties, due to the restriction to a one-parameter model, there is not much flexibility in the model. Additionally, the assumption of equidispersion is frequently incorrect in practice.

The Conway-Maxwell-Poisson (CMP) distribution is a two-parameter extension of the Poisson distribution that generalizes some well-known discrete probability distributions including the Poisson, Bernoulli and Geometric distributions. It also leads to the generalization of distributions derived from these discrete distributions such as the binomial and negative binomial distributions (Shmueli et al., 2005). This distribution, originally introduced by Conway and Maxwell (1962), has subsequently been neglected or forgotten for decades and later on revived by Shmueli et al (2005).

The CMP distribution consists of an extra parameter next to the rate parameter of the Poisson distribution, which we denote by v, and which governs the rate of decay of successive ratios of probabilities such that (Shmueli et al., 2005):

$$\frac{P(X = x - 1)}{P(X = x)} = \frac{x^{v}}{\lambda}$$

This extra parameter can add flexibility to the model to account for the over- or underdispersion of the data which has often been a practical problem for the Poisson distribution. Furthermore, it allows for a variety of generalizations such as zero-inflated data and dependence. The CMP is like the Poisson a member of the exponential family with probability distribution:

$$P(X = x) = \frac{\lambda^{x}}{x!^{v}} \frac{1}{Z(\lambda, v)},$$

where $Z(\lambda, v) = \sum_{i=0}^{\infty} \frac{\lambda^{i}}{j!^{i}}$ is a normalizing constant such that the probability mass function sums up to 1.

If v equals 1, the resulting distribution equals a Poisson distribution with rate parameter λ.

Kokonendji et al. (2008) exhibited that when v ∈ [0,1], the data was over-dispersed and under-dispersed when v ∈ [1,∞].

The ease of use, its flexibility towards over and under-dispersed data and the fact that it generalizes other important discrete distributions makes it appealing for modelling scoring rates in soccer. Moreover, the CMP distribution is new to statistical modelling of football outcomes.

3) *Discrete Weibull distribution*

The Poisson distribution has been widely used due to its attractive mathematical properties. Additionally, it is one of the few count models that describes time between events, which also explains its popularity. However, the Poisson also implies that the inter-arrival times of the model have a constant hazard rate with respect to time. In soccer, the inter-arrival times represent the time between goals, and depend on various factors such as fluctuations in performance, weather conditions etc. , which is counterintuitive to the hazard rate of the Poisson distribution which is constant.

The continuous Weibull distribution has the characteristic that the hazard rate is proportional to time, which ensures that the inter-arrival times can vary over time.

McShane et al. (2008) use Taylor series expansion of the exponential in the Weibull density to derive a discrete version of the Weibull probability distribution where the inter-arrival times are assumed to be independent and identically distributed Weibull random variables. Later on, Bolshakov et al. (2016) introduced this discrete form of the Weibull distribution to the world of soccer statistics, which is in turn more flexible than the initial proposed Poisson model. The probability mass function of the Weibull count distribution is the following:

$$P(X(t)=x)=\sum_{j=x}^{\infty}\left(\frac{(-1)^{x+j}(\lambda t^{c})^{j}\,\alpha_j^x}{\Gamma(cj+1)}\right),$$

where c represents the shape parameter and λ the scale parameter of the distribution, which can alter the shape of the distribution allowing for more flexibility.

4) *Zero-Inflated Poisson*

Besides the frequent violation of the equidispersion assumption, most count data expresses a larger amount of zeros than the Poisson can handle. Zero-Inflated Poisson distributions can offer aid to solve this problem (Lambert, 1992). It is mixture distribution with two different components. The binomial component aims to model the probability of excess zeros and the Poisson component accounts for the non-excess zeros and the non-zero counts (Loeys et al., 2011). The Zero-Inflated Poisson model can be seen as a mixture of both components and translates into the following probability distribution:

$$(1)\ \ P(X=0) = p_i + \left(1-p_i\right) e^{-\lambda}$$

$$(2)\ \ P(X=k) = \left(1-p_i\right) e^{-\lambda} \frac{\lambda^k}{k!}$$

The two components denote (1) the probability of an excess zero and (2) the probability for the non-excess zeros.

2.4 Modelling the correlation between teams

If two random variables are assumed to be independent, the joint distribution results in the product of the two marginal distributions. In case of dependence between the random variables, the product of these marginals is no longer valid and more complicated calculations are necessary to form a joint distribution.

The basic model as suggested by Maher et al. (1982) assumes independence between the goal scoring abilities. In practice, this relates to two teams that score and prevent to score on different pitches without taking into account the qualities of the opponent, speed of the game etc. (Maher et al., 1982). A correlation between the scoring abilities is thus both theoretically as practically valid. Maher et al. (1982) were the first to suggest that a small correlation could give a significant improvement over the independent model. Karlis et al. (2003) pursued the idea of a correlation between the two distributions by using a bivariate Poisson model with an extra parameter to account for the correlation. Bolshakov et al. (2016) were the first to suggest a copula to construct a bivariate model.

Copulas provide a way to model the marginal distributions and its dependence structure separately. A copula is defined as a function which is a cumulative distribution function where the marginals are uniformly distributed. A copula is often used to describe the dependence between random variables (Schmidt et al., 2006). Intuitively, the marginal distributions are transformed into uniform ones and the copula then expresses the dependence structure between those marginals (Schmidt et al., 2006). Most probability distribution functions can therefore be used to model joint probability distributions.

Two random variables X_1 and X_2 are taken as an example, with their associated cumulative distribution functions $F(X_1)$ and $F(X_2)$. The latter are also denoted as the marginal distributions. The dependence structure may be expressed setting $u_i = F(X_1)$ via

$$C\left(u_1, u_2\right)$$

This function is known as a copula and describes the dependence structure separated from the marginals (Schmidt, 2006).

Considering the ease of use, the choice was made to use copulas to model the different probability distributions with their dependence structures. The choice of copula was restricted to the parametric ones and estimation was done with the R packages Copula and VineCopula.

2.5 Different adjustments as proposed in the literature

1) Dixon-Coles Time Weighting function

Although historical data are indispensable, the maximum likelihood estimation suggests a static nature of team performance while performance of a team tends to fluctuate over time due to a multitude of factors such as managerial changes, transfers, form of a team, etc.
Dixon et al. (1997) addressed this static nature of the parameters and constructed a weighted likelihood were recent performances have higher weights than performances in the past.
The weighting function $\phi$is defined as follows:

$$\phi(t) = e^{-\xi t}$$

as such that the previous match results get down weighted exponentially according to the parameter ξ which nests the static model if ξ= 0.

The combination of the weighting function ϕ with the maximum likelihood equation of section 2.2 results in the following pseudo-likelihood:

$$L(\alpha,\beta,\eta,i = 1,\ldots,n) = \sum \left(\phi\left(t - t_k\right) \cdot \left(-\lambda_h + x_i \cdot \log_e\left(\lambda_h\right) - \log_e\left(x_i!\right) - \lambda_a + y_i \cdot \log_e\left(\lambda_a\right) - \log_e\left(y_i!\right) \right) \right)$$

where t_k is the time match k is played, λ_h and λ_a are the mean goal scoring intensities and $x_i$and $y_i$are the observed goals in match i.

Maximizing the likelihood at time t gives parameter estimates which are based up to time t. In this way, the model has the capacity to reflect changes in team performance (Dixon et al., 1997). Moreover, the historical data gets down weighted to a lower or higher degree depending on the choice of ξ (Dixon et al., 1997), i.e. the higher the value of ξ, the more weight is given to recent results in comparison to past performances.

Comparison of log-likelihoods cannot be used for optimization of ξ due to the fact that it includes a sequence of non-independent log-likelihoods (Dixon et al., 1997). A predictive profile log-likelihood to maximize ξ is suggested to overcome this issue. Dixon et al. (1997) argue that the choice for ξ should be the one that maximizes the predictive capability of the model, focussing at correctly predicting match outcomes instead of score lines. The probability of a home win, a draw and an away win are calculated as follows:

$$p_k^H = \sum_{l,\, m} \left(P\left(X_k = l, Y_k = m \right) \right) \qquad where\ l > m$$

$$p_k^D = \sum_{l,m} \left(P\left(X_k = l, Y_k = m \right) \right) \qquad where\ l = m$$

$$p_k^A = \sum_{l,m} \left(P\left(X_k = l, Y_k = m \right) \right) \qquad where\ l < m$$

The probability of a home win, a draw or an away win for a certain match k are represented by p_k^H, p_k^D and p_k^A. A predictive profile log-likelihood is constructed with the former definitions for the different probabilities with

$$S(\xi) = \sum_{k=1}^{N} \left(\delta_k^H \log_e\left(p_k^H\right) + \delta_k^D \log_e\left(p_k^D\right) + \delta_k^A \log_e\left(p_k^A\right) \right),$$

where δ_k^H is 1 if the match results in a home win and 0 otherwise. The same goes for δ_k^D and δ_k^A. Subsequently, a predictive profile log-likelihood plot of S(ξ) against$\xi$is constructed to visualize where the maximum of ξcan be found. Time units are expressed in days instead of half weeks as proposed by Dixon et al. (1997) with the reasoning that results from the day before may influence other fixtures. Moreover, a lot of matches are postponed during the season because of the crowded game calendar for some teams who compete in multiple competitions. Five subsequent seasons plus the first 10 games of the next season were used as training data. The following games of the remaining 28 fixtures were calculated with inclusion of the last game to estimate the next. The 280 predicted matches were used afterwards to examine the maximization of ξ. This procedure was repeated five times in order to have specific values of ξ for every season that was predicted during this analysis. This estimation procedure was performed for the Poisson distribution both for the independent case as the bivariate with copula model.

2) Dixon-Coles adjustment for low scoring games

The second adjustment proposed by Dixon et al. (1997) is an adjustment for low scoring games. The assumption of independence seems valid when comparing the observed and predicted outcomes with the exception of the four low scoring games, which are the 0-0, 1-0, 0-1 and 1-1 match scores (Dixon et al., 1997). The proposed adjustment alters the log-likelihood equation with the estimation of the extra parameter τ and therefore creates some sort of dependence. It is constructed as follows:

$$\tau_{\lambda^H \lambda^A}(x,y) = \begin{cases} 1-\lambda^H \lambda^A \rho & if \;\; x=y=0 \\ 1+\lambda^H \rho & if \;\; x=0, y=1 \\ 1+\lambda^A \rho & if \;\; x=1, y=0 \\ 1-\rho & if \;\; x=y=1 \\ 1 & otherwise \end{cases}$$

The parameter ρ gets maximized in the log-likelihood and corresponds to independence when ρ = 0.

3) Rue adjustment

Rue et al. (1998) tend to factor in the psychological effect of underestimation or overestimation of team performances. The goal scoring abilities are modelled as follows:

$$\log_e(\lambda_h) = \alpha_h + \beta_a - \gamma\Delta_{ha} + \eta$$
$$\log_e(\lambda_a) = \alpha_a + \beta_h + \gamma\Delta_{ha},$$

Where α and β represent the attacking and defensive qualities of the home and away team and η the home advantage parameter. Additionally, Δ_{ha} denotes the difference in strengths between both teams, calculated by subtracting the difference of the sum of attacking and defensive strength per team:

$$\Delta_{ha} = \frac{(\alpha_h + \beta_h - \alpha_a - \beta_a)}{2}$$

Underestimation (or overestimation) is a psychological effect that occurs when the opponent's strength is perceived as a lower value (or higher) than its value in reality. For example, if team h is stronger than team a, there is a possibility that team h would underestimate the abilities of team a. This could result in a more laid back attitude of team h and result in poorer performance of team h.

The γ parameter is a constant modelled to estimate the magnitude of this psychological effect (Rue et al., 1998). A positive γ would imply a psychological underestimating effect, meaning that stronger teams would underestimate weaker teams on average, whereas a negative γ would imply an overestimating effect of the better team.

4) Combination of the time, Dixon Coles and Rue adjustment

Previously, estimation and testing of the adjustments was done separately. However, the literature only tests a combination of adjustments instead of the adjustments separately. Consequently, testing of a combination of the former adjustments was performed. A combined likelihood function was established with ξ as the time weighting function as discussed in 2.4.1, tau the adjustment for low scoring games as discussed in 2.4.2 and the γ parameter to model the psychological effect of over or under estimation as discussed in 2.4.3. This complex likelihood tends to account for the dynamic nature of the game, the underestimation of low scoring games and psychological effects. Additionally, it nests the independent Poisson model if ρ, ξ and γ parameters equal 0.

5) Streak adjustment

A lot of different psychological effects have been hypothesized in sports. In basketball for example, the hot hand fallacy led to a widespread discussion. In this dissertation, the effect of a so called streak on the predictive performance of the different models has been examined.
A streak is a series of won games or gaining a large number of points in the table. The performance of teams fluctuates over the course of a season and the streak variable tends to captivate the short term team performance. This is quantified by using the amount of points acquired in the previous games instead of using their goal scoring abilities. Here fore, the three points for a win system was used, awarding 3 points to the winning team, 0 for the losing team and 1 point for both teams in case of a draw. The purpose of using this point system was to reward the outcome, rather than the estimated strengths of the teams.
The effect of the amount of points won in the period of three matches and five matches has been evaluated. However, due to the varying time spans in which games are played and the sometimes long intermissions between games, short term effect would not be valid anymore. For example international breaks or between season breaks. Therefore, an additional correction for the time period has been taken into account. This results in the following equation:

$$S = \frac{\sum_{i=1}^{n} (P_i)}{D},$$

where P_i is the number of points earned for a certain match i and D denotes the difference in days between the first and last game of the prespecified period. Thus, S relates to the number of points won in a certain time span.
Intuitively, the division of the streak by the number of days is valid due to the fact that good form depends on short time periods. In this way, streaks over long time periods, such as when summer breaks or international competition breaks occur, are down weighted.

2.6 Maximum Likelihood estimation

Maximum likelihood estimation is a popular statistical technique, used in various research domains.

The likelihood function is the function that defines:

$$L(\theta \mid Y) = \prod \left(\Phi\left(Y_i \mid \theta \right) \right)$$

where Y denotes the random variable and ϑ represents the parameter vector for the model.
In all situations the likelihood is the joint density of the observed data to be analyzed and the Maximum Likelihood Estimator (MLE) is the estimator that maximizes the likelihood. Intuitively, the MLE results in the most likely estimator based on the available data. The likelihood is the foundation of classical model-based statistical inference and serves as a natural starting point for frequentist based statistical inference (Boos, Stefanski, 2013).
MLE was used to estimate the goal scoring probabilities as explained above.
All computations were performed in RStudio v 1.1.463. Maximization of the MLE was estimated with the optim package and the nlm package. These packages cover both quasi Newton methods and Newton type methods with optional box constraints for optimization.

A combination of all the adjustments, probability distribution functions and the use of a copula gave a total of 20 models for the comparison of predictive quality. All Poisson based distribution functions were evaluated both independently and with a copula. The static time model, the time weighting adjustment and the Rue adjustment were all evaluated separately with every aforementioned distribution with the exclusion of the CMP model. One independent

Poisson model was additionally built with the adjustment for low scoring games and one with a combination of the time function, Rue adjustment and Dixon-Coles adjustment.

2.7 Assessing model fit

2.6.1 In sample performance

In sample performance was evaluated both for the univariate and bivariate distributions by means of log-likelihood and AIC values. Graphical inspection was additionally performed.

2.6.2 Out of sample performance: The Ranked Probability Score

The ranked probability score (RPS) is a measure of how good forecasts that are expressed as probability distributions match observed outcomes (Baboota et al., 2019) and was first introduced by Epstein (1969). It is both strictly proper (Murphy, 1969) and sensitive to distance (Murphy, 1970). As both the location and spread of the forecast distribution are taken into account, the RPS is a valid method to compare probabilistic models in the prediction of football outcomes (Constantinou et al., 2012) and is defined as follows:

$$RPS = \frac{1}{r-1}\sum_{i=1}^{r}\left(\sum_{j=1}^{i}(P_j) - \sum_{j=1}^{i}(O_j)\right)^2$$

It represents the sum of the squared differences in cumulative distribution functions of the predictions and the observed outcomes, divided by the number of instances minus 1. P_j stands for the predicted probability of an outcome and O_j stands for the observed outcome, while r denotes the number of possible outcomes, being three in the case of football. The use of cumulative distributions makes sense because forecasting soccer matches have ordered outcomes. This implies that for example a draw is closer to a win than a defeat in football.

RPS outcomes are continuous by nature and result in a value between 0 and 1, where the lower the RPS, the closer the cumulative distribution of the prediction is to that of the outcome, resulting in better predictive performance.

Values for the RPS are calculated for every match of the test set and for every model separately. Summary statistics will be provided and tests between the mean RPS between the different models are conducted to see if there is a significant difference in predictive power between the models.

2.7 Which method evaluates as the best

Due to the skewness in the distribution of the RPS, an ANOVA test was used to test the hypothesis of different means. A Box-cox transformation was adopted to fulfill the normality assumption of the test.

If necessary, post hoc analysis was performed through pairwise comparison with the Paired t-test to see which method(s) is (are) the best.

Chapter 3: Results

3.1 Univariate exploration

The goodness of fit for the distribution of the home and away goals was explored both graphically and by means of log-likelihood calculations and can be found in table 1 and figure 4 and 5. The Poisson model performs the worst with respect to the log-likelihood and the AIC for both the home and away goals where the discrepancy is the highest for the away goals. Graphical inspection shows good fits for most models except the Poisson which differs predominantly for the scores of 0 and 1 of the visiting team, which is in line with observations of Dixon et al (1997). The Zero-Inflated Poisson can handle both over- and underdispersion (Loey, 2011) and generalizes better to the observed distributions. Both the CMP and the Discrete Weibull perform best based on the log-likelihood, with negligible difference between the two. Additionally, the CMP and the discrete Weibull visually show the best fit on the histograms.

Table 1. Univariate log-likelihoods (upper value) and AIC (lower value) of the home and away goals

	Home Goals		Away Goals	
	Log-Likelihood	**AIC**	**Log-Likelihood**	**AIC**
Poisson	**-6046.491**	**12094.98**	**-5445.038**	**10892.08**
CMP	**-6037.488***	**12078.98***	**-5423.949**	**10851.9**
Weibull	**-6037.862**	**12079.72**	**-5423.717***	**10851.43***
ZIP	**-6041.263**	**12086.53**	**-5425.827**	**10855.65**

*: Best log-likelihood

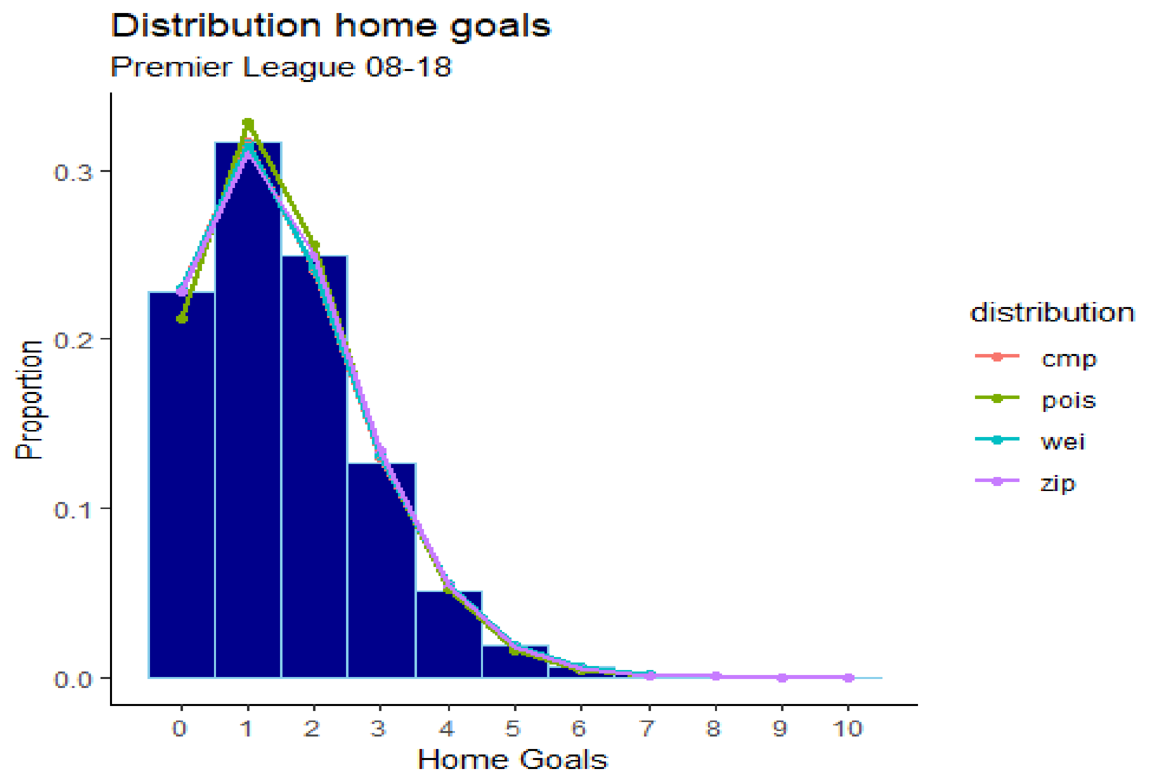

Figure 4. Distribution of home goals

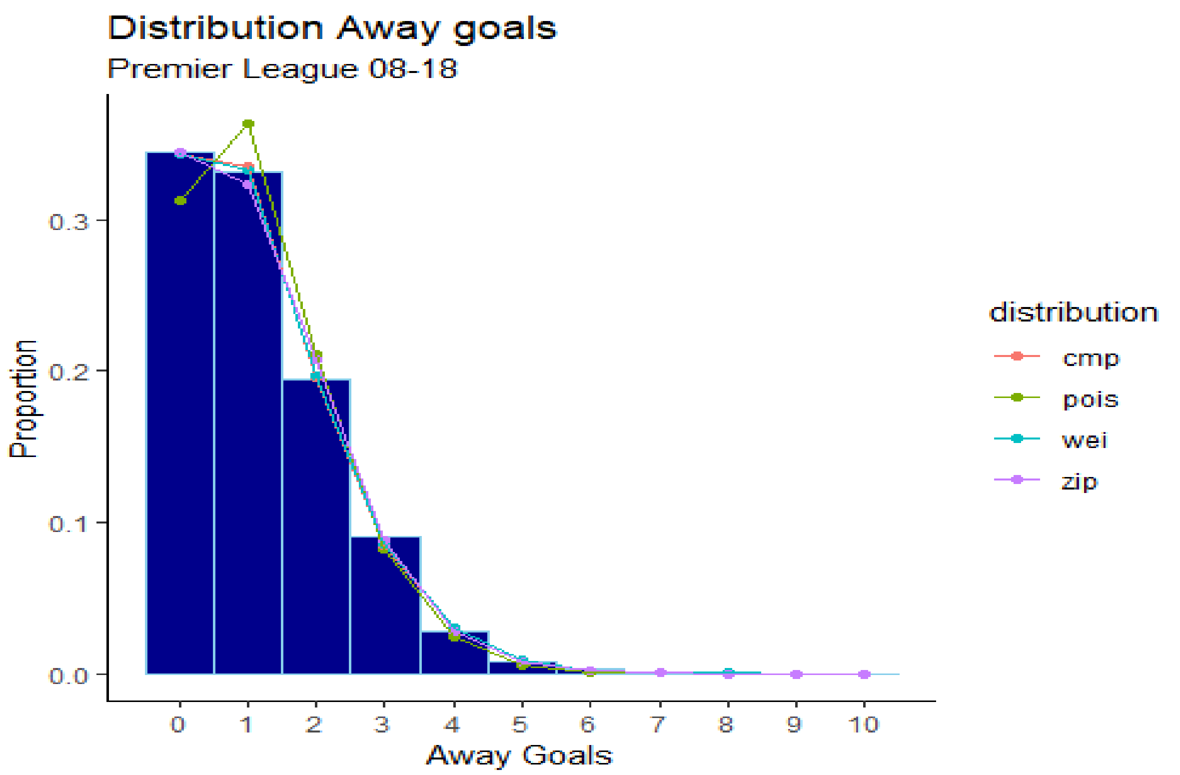

Figure 5. Distribution of the away goals

3.2 Bivariate exploration

Secondly, bivariate properties were examined. Both bivariate distributions modelled with a correlation factor by means of a copula and bivariate distributions without a correlation factor were taken into account (table 2). A small negative correlation was estimated from the data between the home and away goals (-0.071).

Table 2. Bivariate log-likelihoods (upper value) and AIC (lower value)

	Independent		Copula	
	Log-Likelihood	**AIC**	**Log-Likelihood**	**AIC**
Poisson	**-5442.617**	**11001.23**	**-5491.259**	**11098.52**
CMP	**-5438.719***	**10997.44***	**-5487.639***	**11095.28***
Weibull	**-5438.867**	**10997.73**	**/**	**/**
ZIP	**-5442.517**	**11005.03**	**-5503.214**	**11126.43**

The Poisson, Conway-Maxwell Poisson (CMP), the Zero-Inflated Poisson (ZIP) and the Weibull count model were compared based on the log-likelihood and AIC to examine the best fit for the football data. Furthermore, both the independent cases and the models that allow correlation between the goals scored by use of a copula were examined.

Based on the log-likelihood values, the CMP and Weibull count model both perform better than the Poisson model for the independent and copula cases respectively. The ZIP model performs similar for the independent model but is inferior to the Poisson when using a copula. Therefore, the improved performance comes from altering the probability distribution from Poisson or ZIP to CMP or Weibull. Those findings are in line with previous results from Boshnakov et al. (2016). Interesting is the similar fit of the CMP and the Weibull Count model to the data.

The AIC indicates that the CMP and Discrete Weibull perform the best of the four distributions, with a negligibly difference between them both.

Additionally, the bivariate distribution of the observed goals can be found in figure 6.

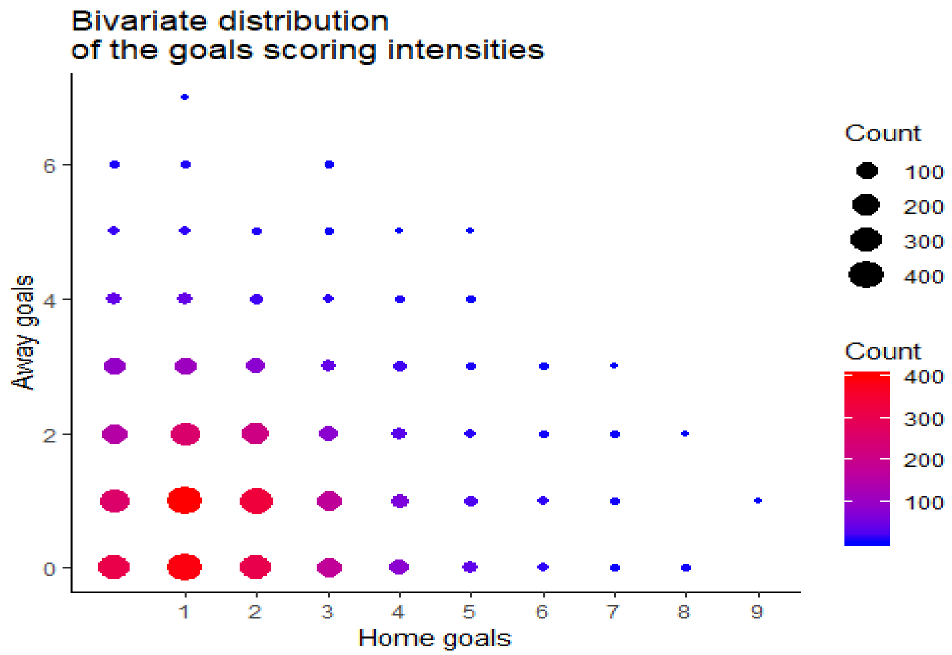

Figure 6. Bivariate distribution of the home and away goals

3.3 Predictive profile log-likelihood

Considering the fact that Dixon et al. (1997) found 0,0065 to be the value to maximize the predictive profile log-likelihood, all values between 0,0001 and 0,0100 with an increment of 0,0001 were used to search for a local optimum for the five seasons prior to our test data. The results are displayed in figure 7.

Dixon et al. (1997) argues that the choice of ξ should be selected in order to optimize the probability of correctly predicting the match outcomes instead of correctly predicting the goals scored by each team as explained in section 2.4. This is in line with the intent of this study: finding the best model to predict match outcomes and is thus used to optimize ξ.

A value of 0.0010 was found as an optimum for the independent models, which is remarkably lower than the outcome of Dixon et al. (1997), translating into a lower degree of dependency of recent results. A value of 0 was found to be optimal for the copula models. Therefore, no time adjustment was used for any of the copula models.

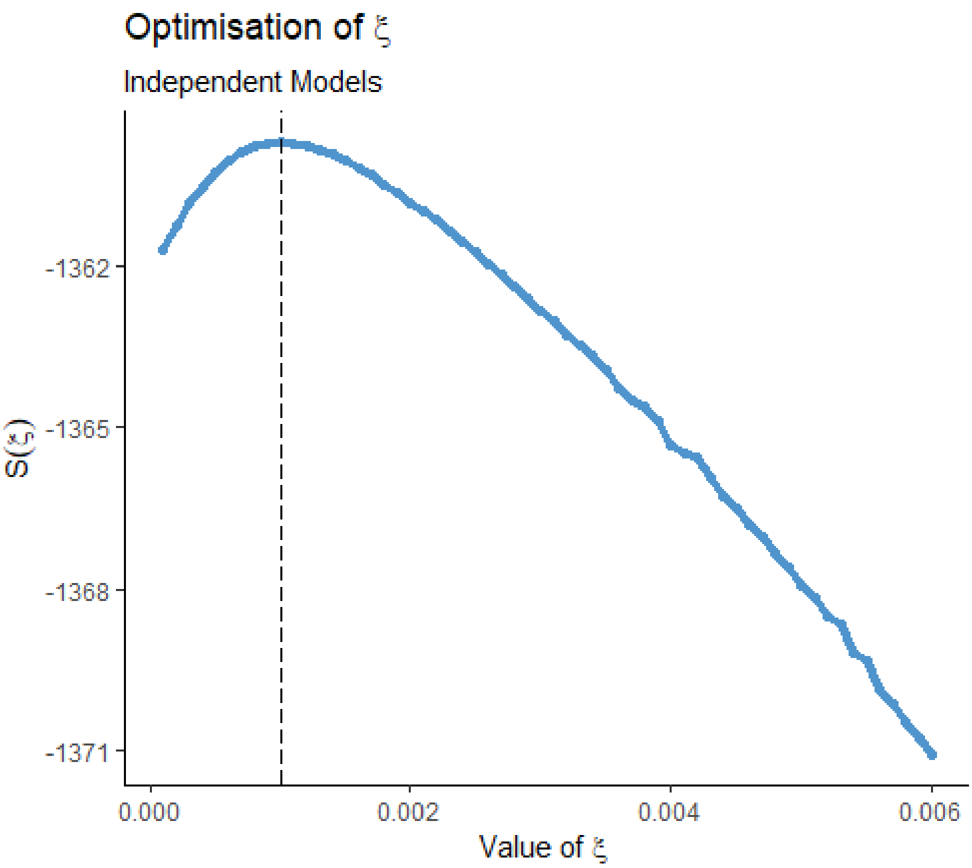

Figure 7. Optimization of the time parameter

3.4 The adjustments

The evolution of the Dixon-Coles adjustment parameter through time

Only the independent Poisson and copula Poisson models seem to have the need for an adjustment for low scoring games. Therefore, solely those distributions are modelled with the Dixon-Coles adjustment. The evolution of the ρ parameter through the course of the estimated five seasons, with exclusion of the first 10 games of each season, is visualized in figure 8.

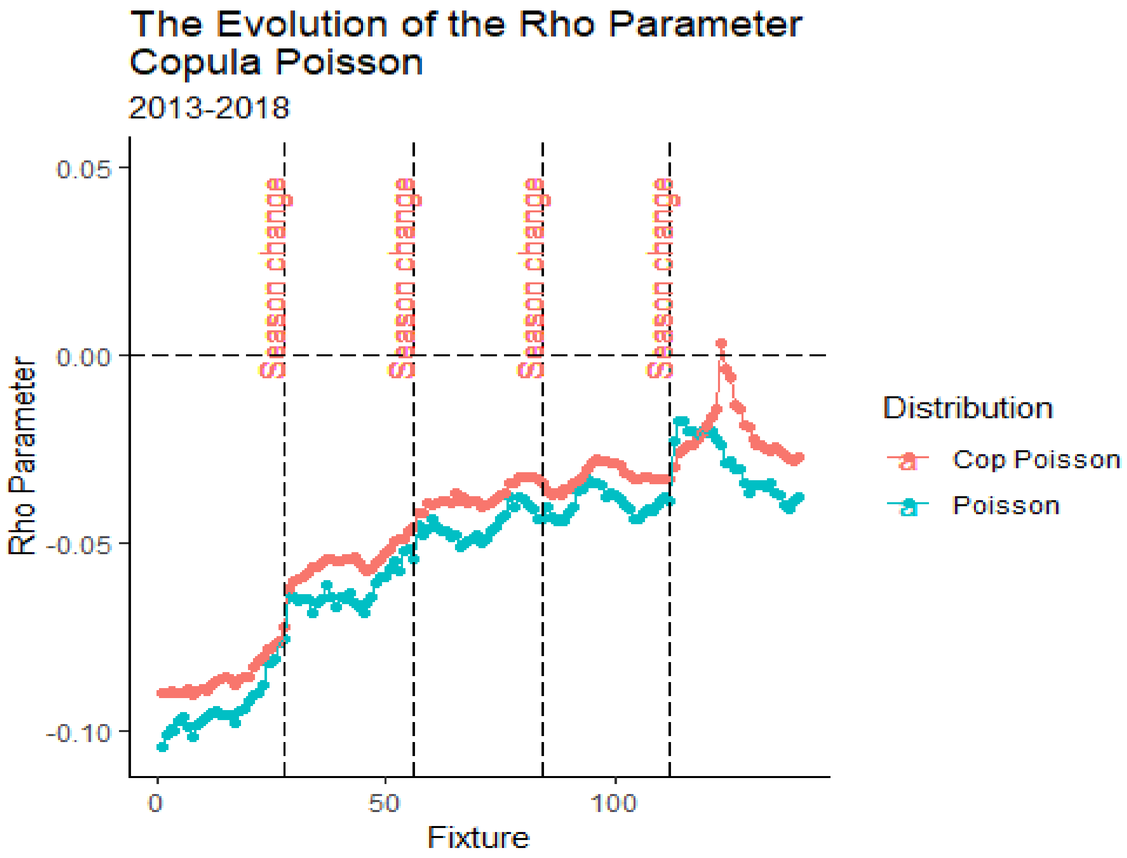

Figure 8: The evolution of the ρ parameter

The ρ parameter tends to be both negative and increasing throughout time for both models. The copula induced model shows a smoother course which implies less match to match variation and crosses the 0 value (which implies independence) in the last season once, implying a decrease in importance of the adjustment over the years. Due to the fact that more efficient models do not need such an adjustment and that the importance decreases for the models that do need it, it can be argued whether this adjustment is still necessary.

Furthermore, a likelihood ratio test was performed to investigate whether there is a significant difference between the independent Poisson model and the independent Poisson model with the Dixon-Coles adjustment. A p-value of 0,023 was found, which states a significant difference and therefore still a valid addition to the basic independent Poisson model.

Although the ρ parameter is in fact valid for modelling, the importance has changed throughout time with a regression towards the independence model. Therefore, research in terms of validity has to be performed in the future.

The evolution of the Rue adjustment parameter through time

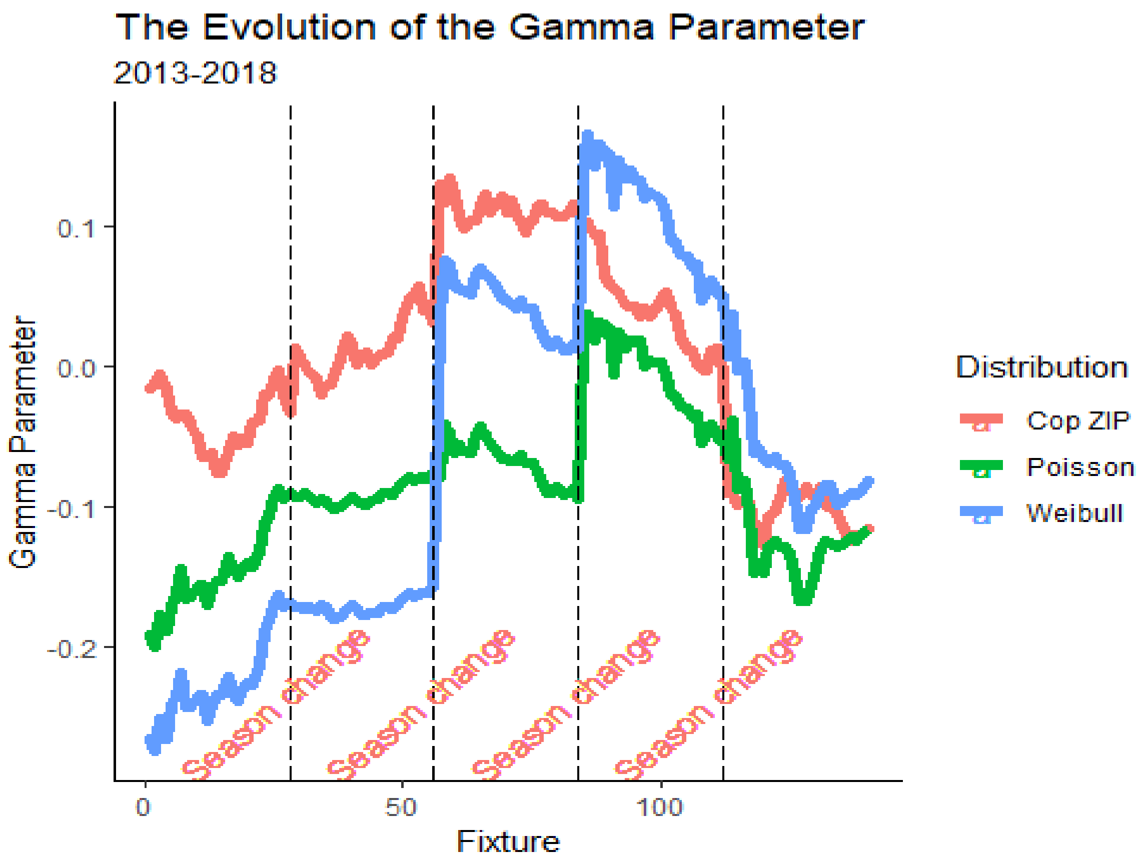

Figure 9: The evolution of the γ parameter

The γ parameter displays a different course. It follows a sinusoidal progress around the 0 value and is subject to large variations from season 13/14 to season 16/17 and an additional drop between season 16/17 and the following season. This pronounced variability is recurrent in both the independent and bivariate models. This high variability may address the minimal effect of the adjustment.

The usefulness of a streak

Both the value of the points earned by a team during the last 3 and 5 matches, with a correction for the time spans were evaluated for every distribution, both independently as with a copula. Results can be found in table 3 and 4.

No significant effect was found for both the 3 match and 5 match based streak variable. Further investigation for predictive capability is therefore abandoned.

Table 3: Results Likelihood Ratio Test (LRT) for the 3 match time span

	Independent		Copula	
	LRT value	**p-value**	**LRT-value**	**p-value**
Poisson	**0.7778321**	**0.377805**	**0.4377479**	**0.5082115**
CMP	**0.7738069**	**0.3790419**	**0.410979**	**0.521473**
Weibull	**0.7789141**	**0.3774734**	**/**	**/**
ZIP	**0.7207868**	**0.3958859**	**0.3838329**	**0.535559**

Table 4: Results Likelihood Ratio Test (LRT) for the 5 match time span

	Independent		Copula	
	LRT value	**p-value**	**LRT-value**	**p-value**
Poisson	**0.4871371**	**0.4852072**	**0.294364**	**0.5874374**
CMP	**0.4859941**	**0.4857197**	**0.2594807**	**0.6104771**
Weibull	**0.5165776**	**0.4723052**	**/**	**/**
ZIP	**0.4829616**	**0.4870839**	**0.2962602**	**0.5862365**

3.5 The Attacking and defensive qualities of teams and the home advantage

Attacking and defensive skills of the Premier League teams were modelled using the CMP distribution with copula as a comprehensive example. Table 5 presents the exponentiated parameter estimates for the last fixture of the 17/18 season.

The higher the attack parameter, the higher the probability of scoring goals and thus the better the attacking performance. The reverse applies for the defensive parameter. The lower the defense parameter is, the lower the probability of scoring a goal for the opponent.

Table 5. Attacking and defensive parameters season 2017/2018 of the Premier League

Team	Attack	Defense	Team	Attack	Defense
Arsenal	**1.643**	**0.750**	**Manchester City**	**1.924**	**0.708**
Bournemouth	**1.185**	**1.120**	**Manchester United**	**1.453**	**0.667**
Brighton and Hove Albion	**0.903**	**0.915**	**Newcastle**	**1.027**	**1.047**
Burnley	**0.871**	**0.855**	**Southampton**	**1.183**	**0.841**
Chelsea	**1.623**	**0.699**	**Stoke City**	**0.992**	**0.933**
Crystal Palace	**1.063**	**0.942**	**Swansea**	**1.081**	**0.975**
Everton	**1.281**	**0.859**	**Tottenham**	**1.593**	**0.754**
Huddersfield	**0.831**	**1.067**	**Watford**	**1.027**	**1.047**
Leicester City	**1.304**	**0.923**	**West Brom Albion**	**0.975**	**0.933**
Liverpool	**1.723**	**0.844**	**West Ham United**	**1.156**	**1.003**

Next, the parameters were examined through time. The evolution of the attacking and defensive abilities and the home advantage were evaluated graphically through the selection of a random team during an arbitrary season. Both the independent models and the dynamic independent models were assessed. The 15/16 Arsenal team was used for this purpose and is subsequently visualized in figure 10.

The parameters appear to be quite different depending on the model. However, they all tend to show the same pattern, a lowering of attacking and defensive qualities during the course of the season. The independent models vary little over the course of a season, while the time weighted models display more variation due to the more weight for recent performances. Of all models, the CMP model with exponential decay seems to vary the most and take the largest decline. Arsenal, a London based club with a rich history, came off at 3rd and 2nd place at the season ending tables for the seasons 14/15 and 15/16 respectively, to end season 16/17 at a 5th place and a 6th place the subsequent season. A decline in both attacking and defensive performance during the span of that season is noticeable but only the time weighted models tend to capture that largeness of decay in performance.

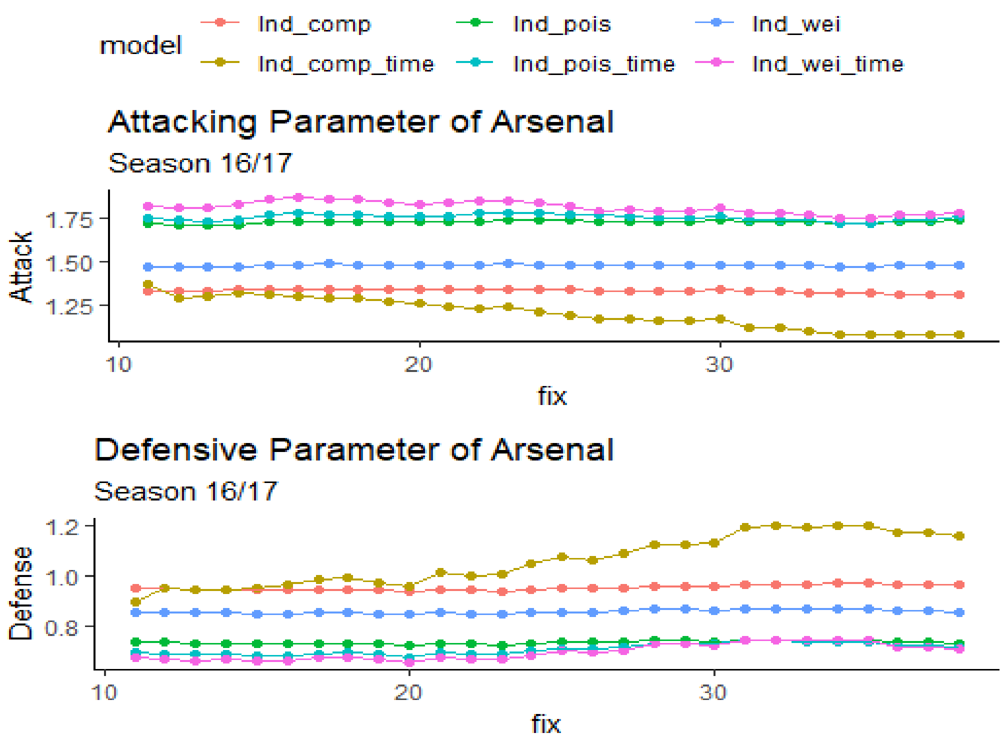

Figure 10. Attacking and defensive parameters of Arsenal during season 16/17

The Home parameter tends to vary similarly as the parameters for the attacking and defensive strengths. The independent models do not let the home parameter vary as much as the time weighted models. This high variability of the home parameter is a bit counterintuitive at first sight. Although the models differ a lot in variability, they again display a similar trend. The benefit of playing at home increases when nearing the end of the season.

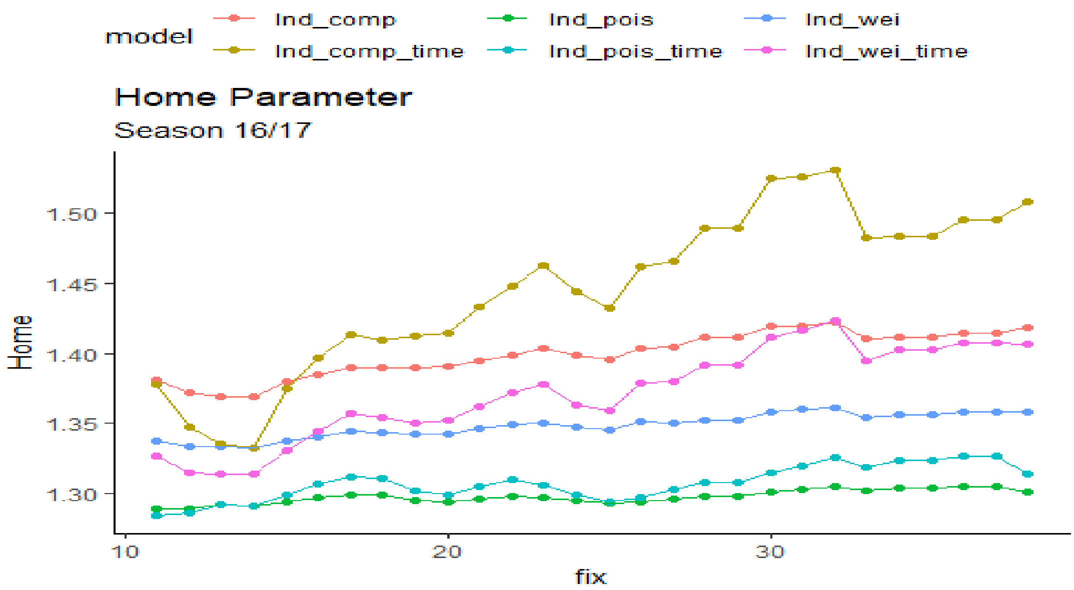

Figure 11. The evolution of the home parameter during season 2016/2017

3.6 Out of sample performance

The ranked probability score was calculated for every predicted outcome as described above, resulting in 1 400 RPS observations per model. Figure 12 represents the distribution of the ranked probability score for the independent Poisson model as an example.

After visual inspection, it was concluded that the distribution of the RPS differed from normality with a skew to the right. Random sampling with replacement by means of the bootstrap with 10 000 repetitions was performed for estimation of the mean of the distribution with respective confidence intervals. Results are presented in figure 13 and table 6.

The CMP with a copula seems to give the best predictions. Although, after transforming the data to a normal distribution, no significant difference between any of the models could be found (p-value of 0.535). Moreover, a Kruskall Wallis test was additionally performed on the non-transformed data, which yielded a non-significant difference as well (p-value of 0.9842).

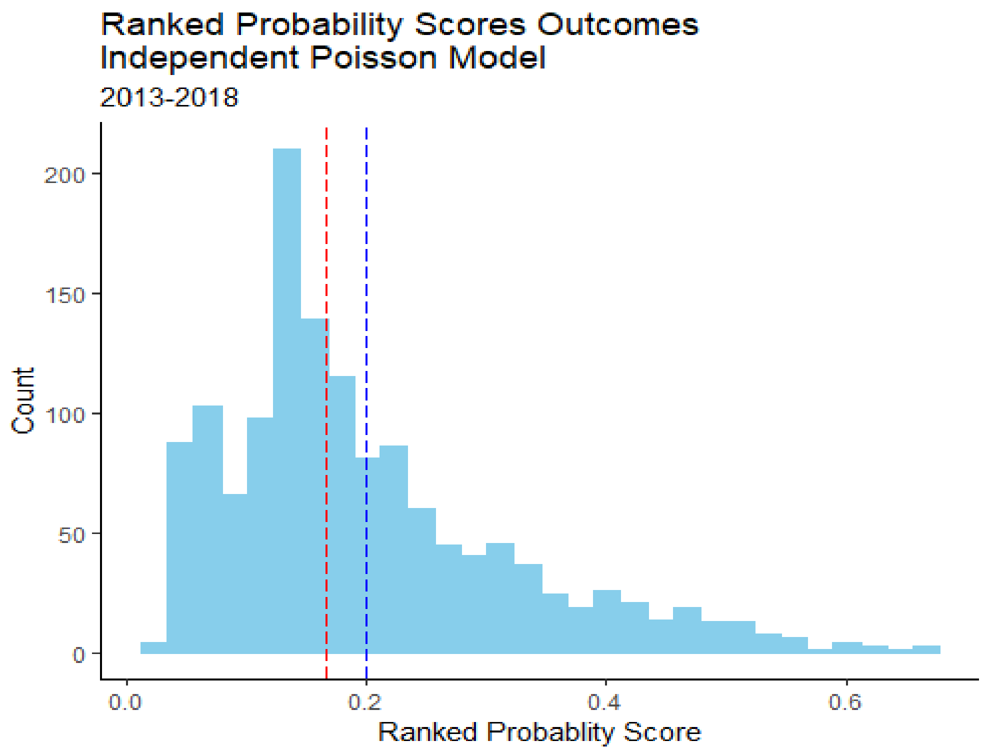

Figure 12. Distribution of the RPS for the independent Poisson model. Red line = Median, blue line = Mean

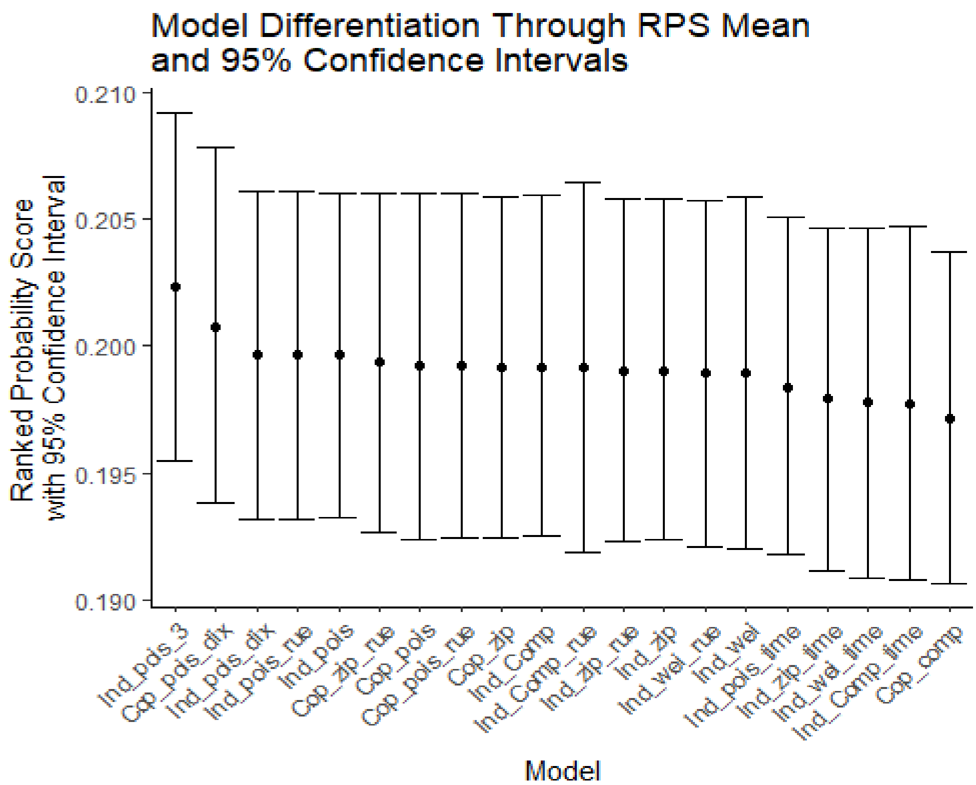

Figure 13. RPS mean and 95% bootstrapped confidence intervals

Table 6. Mean RPS for the different models in the analysis

Model	Mean RPS
Independent Poisson	0.1996339
Independent Poisson with exponential time decay	0.1983938
Independent Poisson with Dixon-Coles adjustment	0.1996554
Independent Poisson with Rue adjustment	0.1996303
Independent Poisson with the 3 adjustments	0.2023569
Zero-Inflated Poisson	0.1990219
Zero-Inflated Poisson with exponential time decay	0.1979495
Zero-inflated Poisson with Rue adjustment	0.199025
CMP	0.1991779
CMP with exponential time decay	0.197709
Discrete Weibull	0.1989191
Conway-Maxwell with Rue adjustment	0.1991339
Discrete Weibull with exponential time decay	0.1989222
Discrete Weibull with Rue adjustment	0.1989222
Copula Poisson	0.1992067
Copula Poisson with Dixon-Coles adjustment	0.1984886
Copula Poisson with Rue adjustment	0.199197
Copula Zero-Inflated Poisson	0.1991845
Copula ZIP with Rue adjustment	0.1992465
Copula CMP	0.1971491

Chapter 4: Discussion and practical examples

4.1 Discussion

The Conway Maxwell Poisson model with the use of a copula was for this dataset the best model of the analysis by means of both in sample performance and predictive capability. However, despite yielding the best predictions, no significant difference could be found between the different models.

The adjustments proposed by the literature are debatable. The Rue adjustment showed to be utmost volatile in time. Furthermore, adjusting for low scoring games is still significant to add to the model, although the course through time suggests it probably will be different in the future. In addition to the proposed adjustments by the literature, the role of a streak has been investigated. No significant effect was found for the streak variable with any of the distributions.

Lastly, putting some dynamics into the independent models improves predictions, while for the copula models it does not.

It should be noted that football itself is a very dynamic sport. Throughout the sport's history, a lot of changes to the game. These changes were made both by the FA (The Football Association), by means of changes in rules, and the clubs and coaches themselves, through the uprising of tactical importance and professionalism. Due to the explosion of the economic value of football and the upcoming data revolution, there will probably be more changes in the way football is played and so, other factors may become more important than they were in the past and vice versa. A practical example is the decay of the ρ parameter through time as found above.

This study only examined the different maximum likelihood based approaches for modelling the goals scored per team as proposed by the literature. Other approaches such as state space modelling (Koopman et al., 2015), Bayesian statistics (Rue et al., 1998), machine learning and Bradley-Terry based models and their modifications were not taken into account.

Furthermore, only free data in the form of the number of goals scored and the times of the matches were taken into account as variables. To get better results, the inclusion of match and player specific variables and confounders can additionally be examined.

The relative low amount of data for newly promoted teams may cause some variability in the estimates, especially those of the newly promoted teams. Further research of this hypothesis is warranted.
A possible improvement would be to include lower tiers of the English football system and cup matches to the data in order to get some interaction between the different leagues. Furthermore, additional variables could also improve the estimates of those promoted teams. Another approach would be to estimate the strengths of the newly promoted teams by their teams value at the beginning of the season.

This dissertation investigated the overall performance of the models using the RPS. Another popular approach for the out of sample performance is comparing the probabilities with the probabilities as proposed by betting firms and to address the predictive power of the model with the expected return on waging on soccer matches, taken the bookmakers odds as a golden standard.

Information in bookmaker’s odds as suggested by Zeileis et al. (2018) can additionally be used in modelling the outcome probabilities. Another addition to the model could be the inclusion of match importance. For example, matches at the end of the season where teams do not fight against relegation or are not able to qualify for European football or to win the competition or cannot rise or fall by position in the table and vice versa, can be quantified by match importance. Furthermore, international tournament matches should be assessed in the model as well. Two reasons may imply why precisely: (1) the match importance factor and (2) seasonal match load may affect team performance due to fatigue or other factors.

4.2 Conclusion

Although the Conway-Maxwell-Poisson model with a copula had the best predictive performance, no significant difference was found between any of the models.
Most adjustments proposed by the literature showed minimal to no effect in predicting the next match outcome. Additionally, the parameters tend to fluctuate heavily over time.

Further research should address the incorporation of variables to the models for both inferential and predictive motives.

Chapter 5: Practical uses of the models:

5.1 Attacking and defensive strengths of teams throughout the course of a season(or longer).

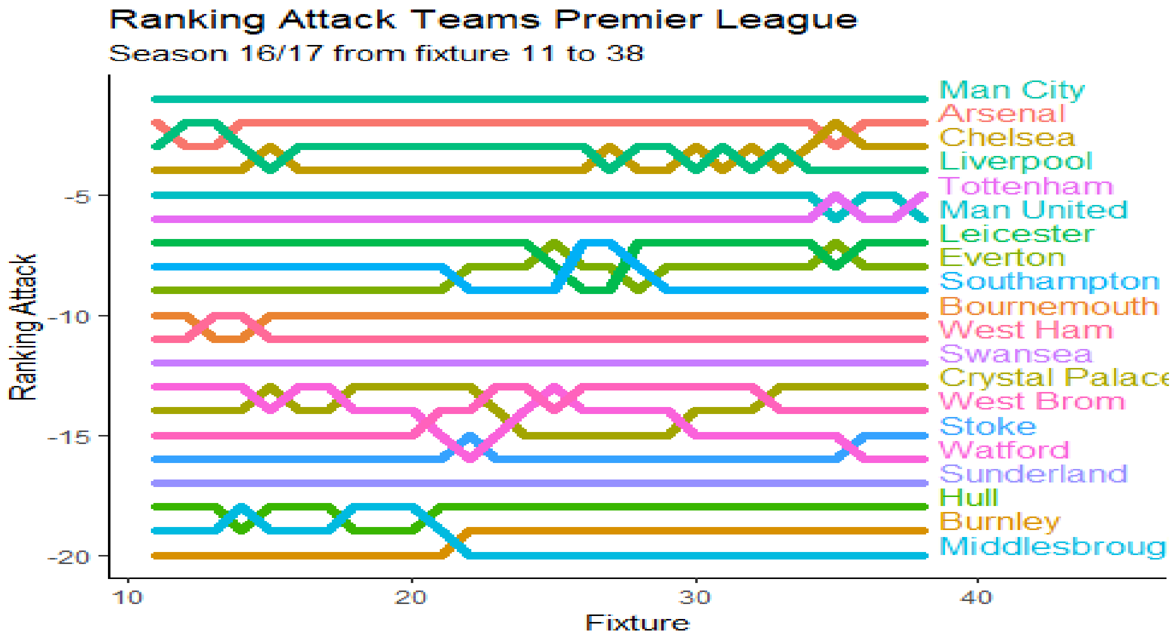

Figure 14. Ranking attacking strengths during season 2016/2017

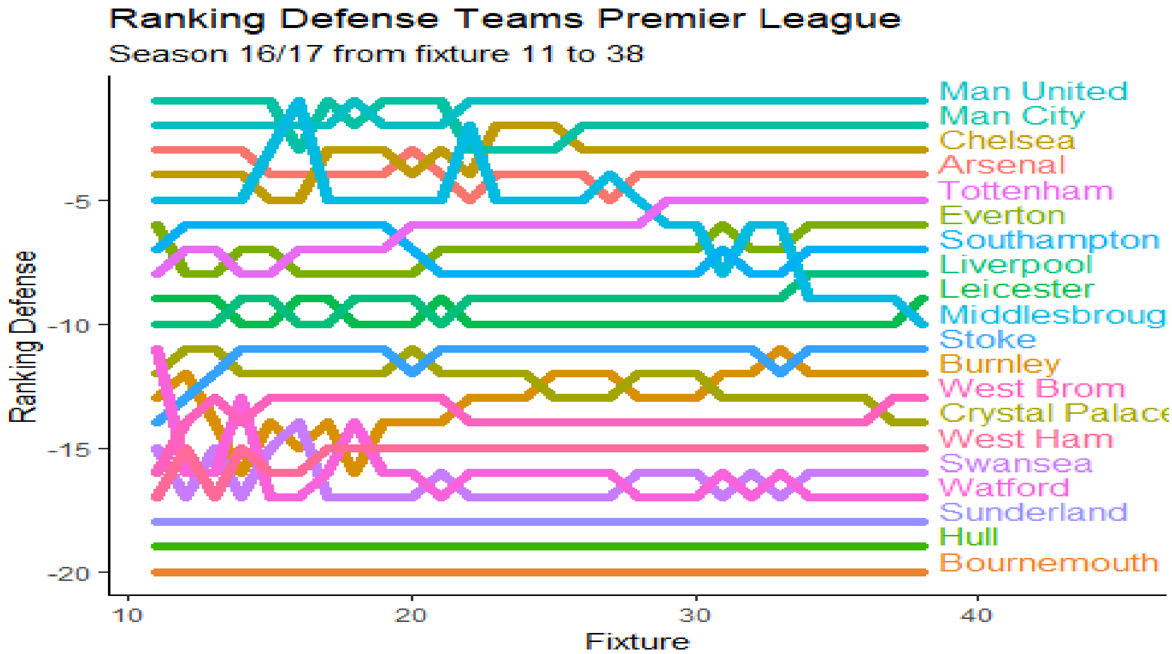

Figure 15. Ranking defensive strengths during season 2016/2017

Figure 14 and 15 display the change in attacking and defensive strengths of the teams of the Premier League during season 16/17 for the Conway-Maxwell-Poisson model with a copula. That season, Sunderland, Hull City and Middlesbrough were relegated from the Premier League. Sunderland and Hull City display poor defensive ratings throughout the entire season, while Middlesbrough lacked attacking strength. Watford and Burnley are also among the lower ranking teams. Watford was able to stay in the league with a peak in performance, both attacking and defending, in the middle of the season, before the wearing off period at the end of the season. Burnley was able to avoid relegation with the aid of its mediocre defense.
The top two finishers, Chelsea and Tottenham respectively, finish worse than the final league table would suggest. Furthermore, Manchester City seems to be the best team overall and the rivals from the other part of Manchester are one of the top defending teams. Fast forwarding to the following season, these are the top two finishers of the subsequent season, with the ‘Citizens’ as clear champions.

5.2 Evaluation of transfer policy and management changes on team performance on both the long and the short term.

Jurgen Klopp has been known to be a fan of ‘gegenpressing’ which he also employed while being the manager of Borussia Dortmund. It is defined as a tactical plan where the team, after losing possession of the ball, immediately attempts to win back possession rather than falling back to regroup (Wiktionary). This style of play is fairly new to soccer and may need some time for the players to adjust to. Adding thereto, Koopman et al. (2015) suggested that most changes in team performance in football are in the long term rather than the short term. The changes in attacking and defensive strengths in figure 16 may aid the hypothesis that teams are built, and organizational decisions such as transfers and choice of trainer may have an impact on the long term. Figure 16 displays the attacking and defensive strengths of Liverpool during the “Jurgen Klopp” -era from the season 15/16 until 17/18. Additionally, important incoming transfers which would become starting players and important factors in the team are displayed. The model can therefore be used to see how teams evolve in the long term. For example, even though there are some short term fluctuations, a long term improvement of attacking and defensive performance is visible since the assignment of Klopp as the manager of Liverpool. This brought them from championship underdog to Champions League finalist in 2018 and title contender in the subsequent season, which is in line with the rise of the team’s quality both on the attacking and defensive end over time.

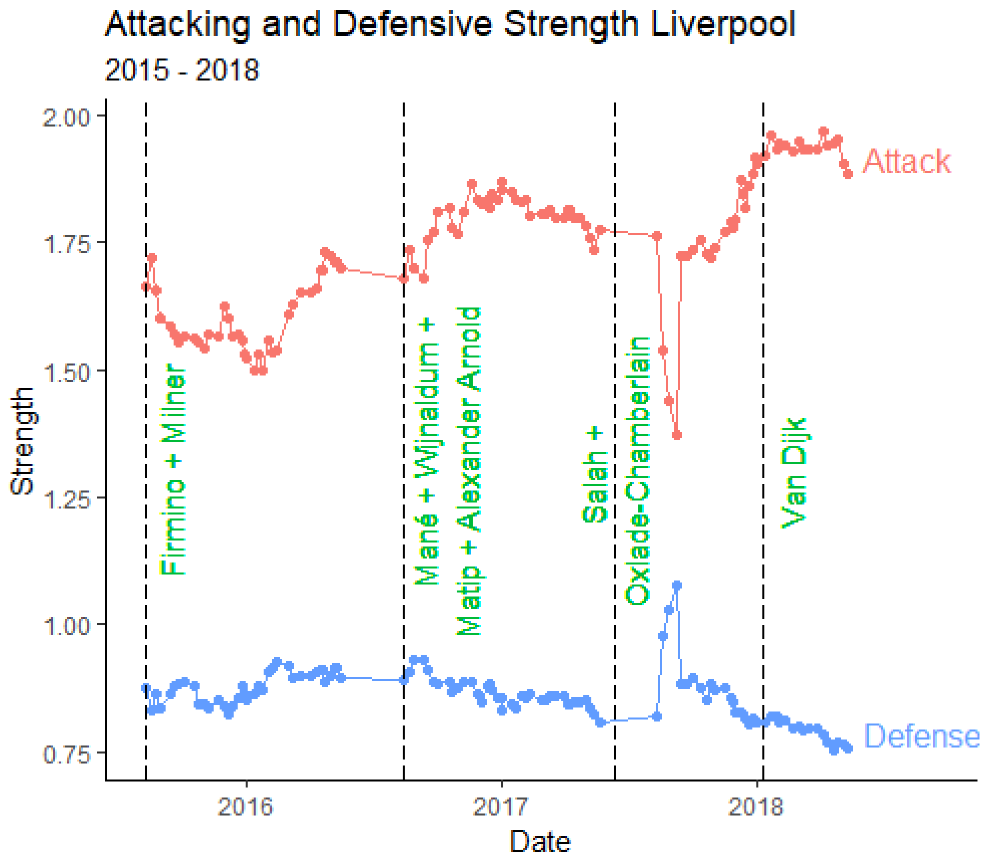

Figure 16. Attacking and defensive strength of Liverpool 15-18

5.3 Forecasting probabilities for the next game

After modelling the data, one can create the probability of a score line for a matchup between two teams. The following explanation is accompanied by a random example. This is the matchup between Tottenham - West Ham United on the 19th of November 2016.

First, the probabilities of each team to score a certain number of goals is calculated based on the model. Combining those two vectors by multiplication results in a matrix with all the possible score lines with their estimated probability. Table 7 yields the estimated score line matrix for the Tottenham - West Ham Game.

Table 7. Estimated goal scoring probability matrix

	Goals Away Team (West Ham United)							
Goals Home Team (Tottenham)		0	1	2	3	4	5	6
	0	0.00283	0.00583	0.00578	0.00374	0.00180	0.00068	0.00021
	1	0.01224	0.02524	0.02502	0.01621	0.00778	0.00296	0.00093
	2	0.02399	0.04948	0.04904	0.03177	0.01524	0.00580	0.00182
	3	0.02984	0.06154	0.06100	0.03952	0.01896	0.00721	0.00227
	4	0.02697	0.05563	0.05514	0.03573	0.01714	0.00652	0.00205
	5	0.01907	0.03934	0.03900	0.02526	0.01212	0.00461	0.00145
	6	0.01105	0.02279	0.02259	0.01463	0.00702	0.00267	0.00084

The elements in the green area symbolize the probabilities where Tottenham scores at least one goal more than the opponent while the red area of the matrix illustrates the probabilities of a score line in favor of West Ham United. The blue boxes represent the probability for a draw between the two teams. Summing those elements of the matrix together gives us the probability of a draw between the two teams. Likewise, summing up the elements in the green area results in the probability of a Tottenham win. The sum of the remaining red part of the matrix results in the probability of a win in favor of West Ham United. Figure 17 illustrates such a summation of the different parts of the score line matrix for a clearer view of the different outcome probabilities.

This is a mere example and can be done subsequently for every game after estimation of the different attacking and defensive capabilities of the teams.

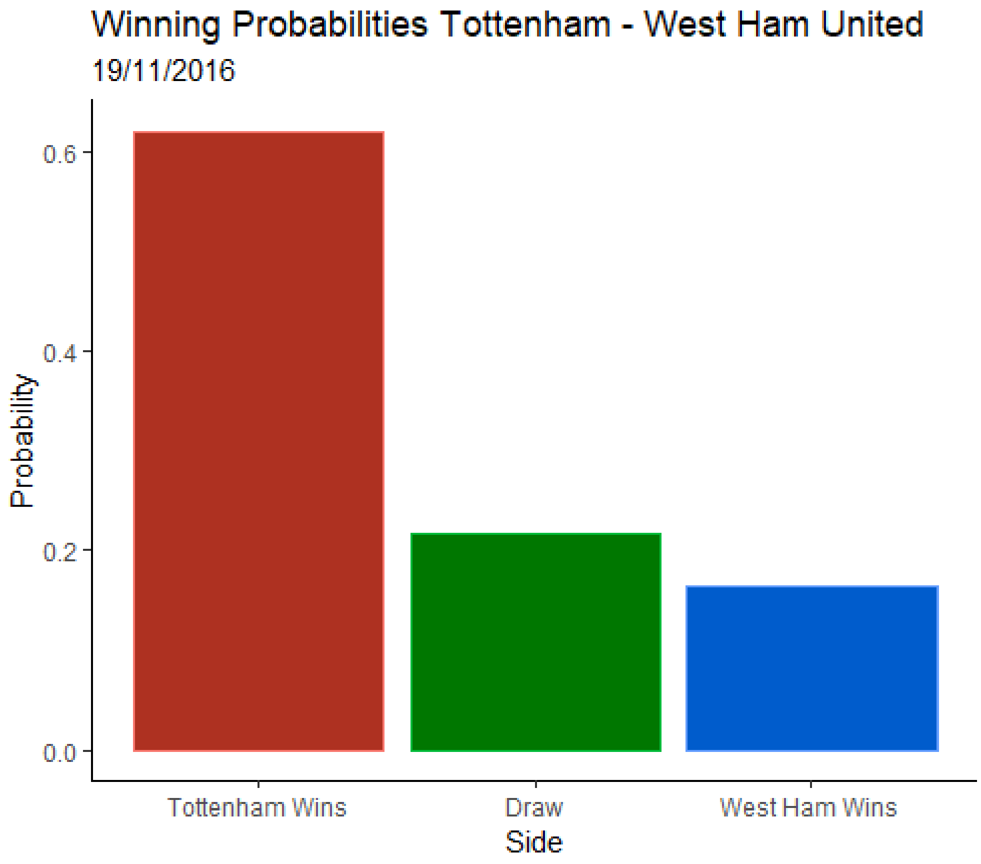

Figure 17. Outcome probabilities for the Tottenham - West Ham United game

5.4 Betting strategy

An alternative way of assessing out of sample performance of the models is by evaluating the returns of a betting regime. The intuition behind this strategy is that the odds offered by the bookmakers serve as golden standard, where bookies are assumed to have far more data available, and are therefore able to build better models that yield better predictive performance. Better performance than the bookmaker would subsequently imply a valid model. However, this hypothesis is flawed in multiple ways. First, betting odds are made in order to maximize returns for the bookmakers, not to impose the best probabilities for each team. Central to this method lies the wisdom of the crowds effect which is one of the methods used to alter betting odds. These alterations are imposed to secure an overall expected winning for the bookies on the long term.

Secondly, the sum of the probabilities provided by the bookmakers usually do not sum up to 100% but sum up to a higher extent. Everything above the 100% threshold are gains for the

bookmakers. The sum of probabilities is reported to be around 107% on average (Koopman et al., 2015), stating an average expected win of 7% for the bookmakers per random bet.
For these two reasons, the strategy of comparing the model performances to the bookmakers odds was abandoned and instead a quantitative metric such as the RPS was used. However, the possibility of a positive return for a betting strategy was investigated.

In order to make a positive return on the long run, our model performance has to beat the bookmaker's edge. Therefore, two popular strategies are implemented
First, a restriction for quality bets is imposed by means of the following equation:

$$EV(A) = P(A) \; x \; Odds(A) - 1 > t$$

where P(A) is the probability of a certain outcome and Odds(A) are the odds offered by the bookmakers and EV(A) is the expected value of an event where it should surpass some threshold t in order to bet on the event. The threshold t can be chosen according to the degree of conservatism of selecting the betting odds. The higher the choice of t, the higher the probability of a positive return per bet, but the fewer bets to wager on and vice versa.

Second, the Kelly criterion (Kelly, 1956) is used to maximize positive return in the long run.

$$f = \frac{p(b+1) - 1}{b},$$

where:

- f is the fraction of capital used to wager on a certain bet
- p is the probability of a certain outcome
- b is the net winnings of a bet

Optimization of the threshold t is done similar to the optimization of the parameter ξ as discussed in section 2.4. Probability estimates from season 2012/13 until 2017/18 were used in the optimization. A grid search to find the optimal value of t was performed with a value of 0.05 as optimum.
Subsequently, both aforementioned betting strategies were incorporated on the estimates for fixture 11 until 23 to calculate the return over time. Additionally, 95% confidence intervals were calculated with the parametric bootstrap method.

A total of 1 910,967 euro was wagered during that time span with an estimated win of 636,167 euro, which evaluates to a positive return of 33%. Concluding that a positive return from betting is definitely a possibility.

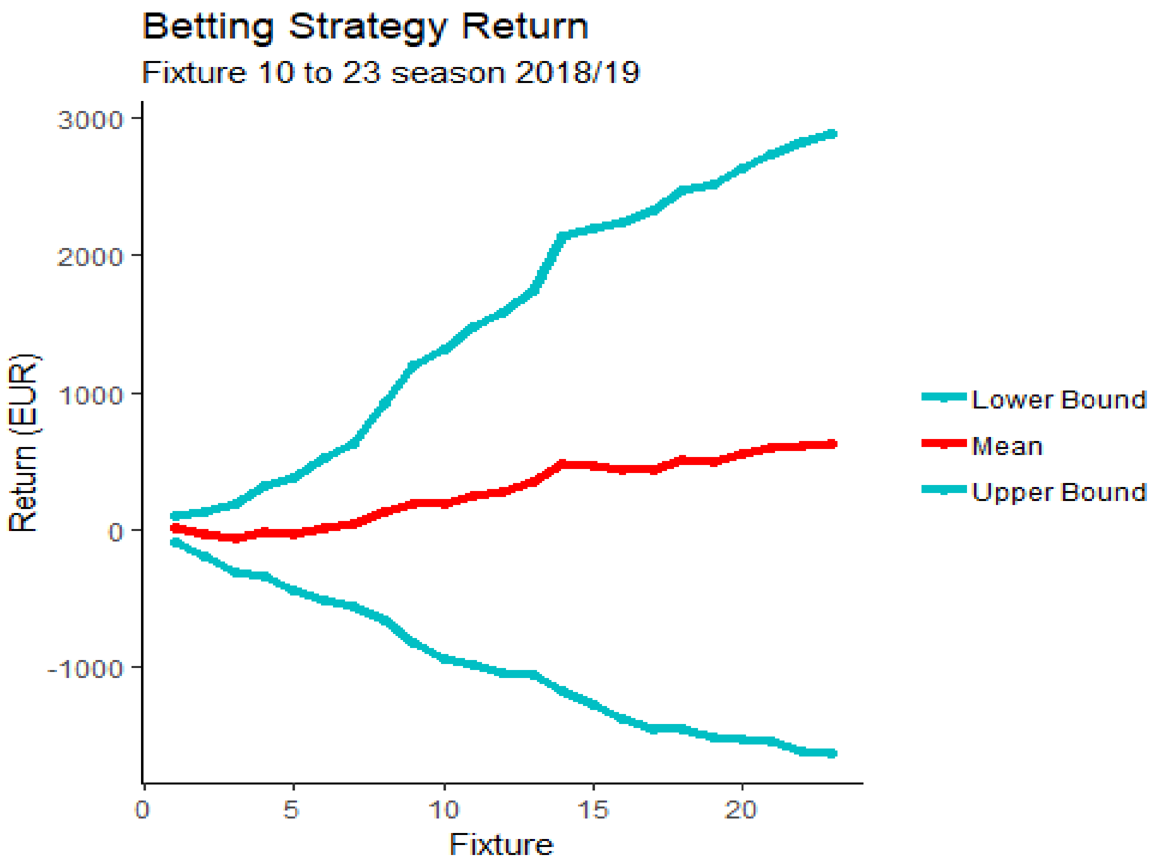

Figure 18. Betting strategy return with 95% confidence intervals

Bibliography

(1) Baboota, R. & Kaur, H. (2019). Predictive analysis and modelling football results using machine learning approach for English Premier League. *International Journal of Forecasting, 35(2),* 741-755.

(2) Boos, D. & Stefanski, L., A. (2013). Essential Statistical Inference: Theory and Methods. Springer.

(3) Boshnakov, G., Kharrat, T. & McHale, I., G. (2016). A Bivariate Weibull Count Model for Forecasting Association Football Scores. *International Journal of Forecasting, 33(2),* 458-466.

(4) CIES, Centre International d'Etude du Sport. Consulted on 12 may 2019, from http://www.football-observatory.com.

(5) Constantinou, A. C. & Fenton, N. E. (2012). Solving the Problem of Inadequate Scoring Rules for Assessing Probabilistic Football Forecast Models. *Journal of Quantitative Analysis in Sport 8(1).*

(6) Dixon, M., J. & Coles, S., G. (1997) Modelling Association Football Scores and Inefficiencies in the Football Betting Market. *Journal of the Royal Statistical Society, Applied Statistics 46(2),* 265-280.

(7) Dunning, E. (1999). The development of soccer as a world game. Sport Matters: Sociological Studies of Sport, Violence and Civilisation. London*: Routledge.*

(8) Epstein, E. (1969). A Scoring System for Probability Forecasts of Ranked Categories. *Journal of Applied Meteorology,8(6),* 985-987.

(9) FIFA Big Count 2006 (2007). *FIFA Communications Division, Information Services.*

(10) Haight, F. (1969). *Handbook of the Poisson Distribution*: Wiley.

(11) Hill, I.D. (1974). Association Football and Statistical Inference. *Journal of the Royal Statistical Society, Applied Statistics, 23,* 203.

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

(13) Kelly, J. L. (1956). A new interpretation of information rate. *Bell System Technical Journal, 35(4),*917-926.

(14) Kokonendji, C. C., Mizere, D. & Balakrishnan, N. (2008). Connections of the Poisson weight function to overdispersion and underdispersion. *Journal of Statistical Planning and Inference, 138(5),* 1287-1296.

(15) Koopman, S. J. & Lit, R. (2015). A dynamic bivariate poisson model for analysing and forecasting match results in the english premier league. *Journal of the Royal Statistical Society: Series A (Statistics in Society), 178(1),*167-186.

(16) Lambert, D. (1992). Zero-Inflated Poisson Regression, With An Application to Defects in Manufacturing. *Technometrics 34(1),*1-14.

(17) Loeys, T., Moerkerke, B., De Smet, O., Buysse, A. (2011) The analysis of zero-inflated count data: beyond zero-inflated poisson regression. *British Journal of Mathematical & Statistical Psychology, 65(1),* 163-180.

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

(19) McShane, B., Adrian, M., Bradlow, E. T., & Fader, P. S. (2008). Count models based on weibull interarrival times. *Journal of Business and Economic Statistics, 26(3),*369-378.

(20) Murphy, A.H. (1969). On the Ranked Probability Score. *Journal of Applied Meteorology, 8,* 988-989.

(21) Murphy, A.H. (1970). The Ranked Probability Score and the Probability Score: A Comparison. *Monthly Weather Review, 98,* 917-924.

(22) Pollard, R. (1986). Home advantage in soccer: A retrospective analysis. *Journal of Sports Sciences 4(3),* 237-48.

(23) R Core Team (2014). R: A Language and Environment for Statistical Computing. *R Foundation for Statistical Computing, Vienna, Austria.*

(24) Reep, C. & Benjamin, B. (1968). Skill and Chance in Association Football. *Journal of the Royal Statistical Society, 131(4),* 581-585.

(25) Rue, H. & Salvesen, O. (1998) Predicting and Retrospective Analysis of Soccer Matches in a League. *Journal of the Royal Statistical Society: Series D (The Statistician), 49,* 399-418.

(26) Schmidt, T. (2006) Coping with Copulas. *Forthcoming in Risk Books "Copulas - From Theory to Applications in Finance.*

(27) Shmueli, G. et al. (2005) A Useful Distribution for Fitting Discrete Data: revival of the Conway-Maxwell-Poisson Distribution. *Journal of the Royal Statistical Society: Applied Statistics., 54(1),* 127–142.

(28) Sklar, A. (1973). Random variables, joint distribution functions, and copulas. *Kybernetika, 9(6),*449-460.

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

(30) Wiktionary. Consulted on 12 may 2019, from https://en.wiktionary.org/wiki/gegenpressing.

(31) Zeileis A, Leitner C, Hornik K (2018). *"Probabilistic Forecasts for the 2018 FIFA World Cup based on the Bookmaker Consensus Model", Working paper 2018-2019, Working Papers in Economics and Statistics, Research Platform Empirical and Experimental Economics, Universtät Innsbruck.*

Modelling Football Outcomes: A Maximum Likelihood Approach

Daan Nyckees

Master dissertation submitted to
obtain the degree of
Master of Statistical Data Analysis

Promoter: Prof. Dr. Christophe Ley
Department of Applied Mathematics, Computer Science and Statistics

Academic year 2018 - 2019