

# A Bivariate Weibull Count Model for Forecasting Association Football Scores

Georgi Boshnakov<sup>1</sup>, Tarak Kharrat<sup>1,2</sup>, and Ian G. McHale<sup>2</sup>

<sup>1</sup>School of Mathematics, University of Manchester, UK.

<sup>2</sup>Centre for Sports Business, Salford Business School, University of Salford, UK.

September 9, 2016

## Abstract

The paper presents a forecasting model for association football scores. The model uses a Weibull-inter-arrival times based count process and a copula to produce a bivariate distribution for the number of goals scored by the home and away teams in a match. We test it against a variety of alternatives, including the simpler Poisson distribution-based model and an independent version of our model. The out-of-sample performance of our methodology is illustrated first using calibration curves and then in a Kelly-type betting strategy that is applied to the pre-match win/draw/loss market and to the over-under 2.5 goals market. The new model provides an improved fit to data compared to previous models and results in positive returns to betting.

**Keywords:** betting, calibration, copula, counting process, soccer, Weibull.

## 1 Introduction

Since the seminal paper by [Maher \(1982\)](#), much effort has been invested in modelling the probability distribution of scores in association football. Maher's model assumes that the numbers of goals scored by each team in a football match follow independent Poisson processes, and that the rates at which the teams can expect to score goals are functions of the ability of the two teams to attack and defend. Subsequent efforts have enhanced the Maher model in a variety of directions. [Dixon and Coles \(1997\)](#) make two enhancements to Maher's model: first, they allow for dependence between the goals scored by the two teams and second, they address the dynamic nature of teams' abilities by using a time-decay function in the likelihood so that more recent results affect a team's estimated strength parameters more than results further in the past. [Rue and Salvesen \(2000\)](#) address the dynamic nature of teams' abilities in a Bayesian framework, as does [Owen \(2011\)](#). [Karlis and Ntzoufras \(2003\)](#) use a bivariate Poisson model with diagonal inflation so that the probabilities of draw scores are better calibrated compared to the simple independent Poisson model. Most recently [Koopman and Lit \(2015\)](#) use a state space model to allow team strengths to vary stochastically with time.

These models all assume the basic scoring pattern in football follows a (time-homogeneous) Poisson process. Perhaps this assumption is made more out of convenience since, other than the negative binomial distribution, there are surprisingly few natural alternatives.

Here, we propose using a count process derived when the inter-arrival times are assumed to follow an independent and identically distributed Weibull distribution. We refer to this model as the Weibull count distribution and, until recently, the form of the distribution for the count process generated by Weibull inter-arrival times was not known. However, [McShane et al. \(2008\)](#) derive this distribution and

so a new, more general, count process model can now be adopted. In addition to using a Weibull count model we allow for dependence between the goals scored by the two teams by employing a copula to generate a bivariate distribution allowing for positive or negative dependence.

Our objective in this paper is to build a model for the goals scored by the two teams in a football match. Our model can be used to construct the probabilities of the score-lines and hence can be employed in betting market analysis and, for example, to study market efficiency.

The computations and the graphs in the paper were done with **R** (R Core Team, 2014). To speed up the computations, where necessary, parts of the code were implemented in C++ using the infrastructure provided by Eddelbuettel and Sanderson (2014).

The remainder of the paper is structured as follows: in Section 2 we present the Weibull count distribution, our bivariate model and give our specification for its use when modelling the goals scored by the two teams in a football match. Results of fitting our model to data from the English Premier League are presented in Section 3 whilst the out-of-sample predictive performance, including the results of a simple Kelly-based betting strategy are described in Section 4. We conclude with some closing remarks in Section 5.

## 2 A Bivariate Weibull Count Distribution

### 2.1 The Weibull Renewal Process

McShane et al. (2008) derive the probability distribution of the number of events occurring by some time  $t$  when the inter-arrival times are assumed to be independent and identically distributed Weibull random variables (this process is also known as a Weibull renewal process). They do so by using a Taylor series expansion of the exponential in the Weibull density. They name the resulting count process the ‘*Weibull count model*’ and its probability mass function is given by

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

where  $\alpha_j^0 = \Gamma(cj + 1)/\Gamma(j + 1)$ ,  $j = 0, 1, 2, \dots$ , and  $\alpha_j^{x+1} = \sum_{m=x}^{j-1} \alpha_m^x \Gamma(cj - cm + 1)/\Gamma(j - m + 1)$ , for  $x = 0, 1, 2, \dots$ , for  $j = x + 1, x + 2, x + 3, \dots$ . In equation (1),  $\lambda$  is a ‘rate’ parameter and  $c$  is the ‘shape’ parameter of the distribution. Here, the observation unit is the match which we take as having a duration of 1 time unit. The rate,  $\lambda$ , is thus the scoring rate per match.

The use of the Weibull distribution to model the inter-arrival times allows the hazard  $h(t)$  associated to the count process to vary over time. The Weibull-hazard is given by

$$h(t) = \lambda c t^{c-1}$$

and can be either monotonically increasing for  $c > 1$ , monotonically decreasing for  $c < 1$ , or constant (and equal to  $\lambda$ ) for  $c = 1$ . Note that when  $c = 1$ , we recover the (time-homogeneous) Poisson process. It is also interesting to note that this model naturally handles both over-dispersed (mean smaller than the variance;  $c < 1$ ) and under-dispersed data (mean larger than the variance;  $c > 1$ ) whilst the Poisson count distribution ( $c = 1$ ) can only accommodate equi-dispersed data (mean equal to the variance).

Despite the somewhat intimidating appearance of equation (1), the computations for the Weibull count model can be done without much trouble. For the usual values of count (goals) observed in association football ( $x \in [0, 10]$ ), the first 50 terms of the infinite series are sufficient to compute accurately

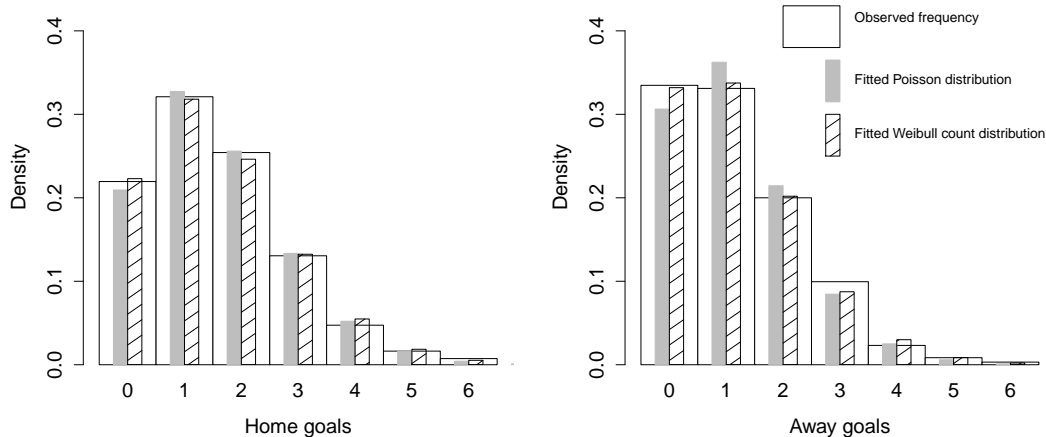


Figure 1: Histograms of home goals (left) and away goals (right) with the fitted Poisson and Weibull count models. The estimated parameters are, for the home team,  $\lambda_H = 1.50$  (0.04),  $c_H = 1.56$  (0.03) and for the away team,  $\lambda_A = 1.10$  (0.03) and  $c_A = 0.85$  (0.04), where the figures in parentheses are standard errors.

the probabilities. For speed, we implemented these computations in C++, though [McShane et al. \(2008\)](#) were able to perform the computations in Microsoft Excel. We validated the computation by retrieving the Poisson case ( $c = 1$ ) and by reproducing the analysis conducted in [McShane et al. \(2008\)](#).

Figure 1 shows the Weibull count model and the Poisson distribution fitted to the goals scored by the home team (left) and the away team (right) in matches played in the English Premier League during the five seasons from 2010-11 to 2014-15. Also shown are the density histograms of home goals and away goals. Eyeballing the fit of the two distributions suggests that the Weibull count model and the Poisson distribution provide similar goodness-of-fit for home goals (although slightly better for the Weibull count especially for the 0 count), whereas for away goals, it is clear that the Weibull count model is an improvement. The  $\chi^2$  goodness-of-fit test statistics for the fitted models shown in Table 1 support this. In fact, they suggest that the Poisson distribution is not adequate for either home or away goals, while the Weibull count model is appropriate.

Table 1:  $\chi^2$  goodness-of-fit test statistics for the fitted Weibull count model and Poisson distribution to home goals and away goals.

	Home Goals			Away Goals		
	$\chi^2$	df	p-value	$\chi^2$	df	p-value
Weibull count model	7.65	4	0.10	6.59	4	0.16
Poisson distribution	13.23	5	0.002	23.5	5	0.0002

We now present a bivariate distribution based on the Weibull count model and include some modifications that can be used for forecasting the results of football matches.

## 2.2 Using a Copula To Generate a Bivariate Model

The existence of some sort of dependence between the goals scored by two teams in a football match is widely accepted. However, the exact specification of the dependence is less clear. For example, [Dixon and Coles \(1997\)](#) find evidence of dependence by studying the difference in the empirical joint

distribution of goals scored by the two teams and the implied joint distribution under the hypothesis that the two random variables are independent (i.e. the product of the marginal distributions). Having confirmed the distributions are not independent, they impose an *ad hoc* correction to their bivariate Poisson distribution. [Karlis and Ntzoufras \(2003\)](#) use a diagonally inflated distribution to account for the fact that draw results are observed more frequently than they would be under the independent bivariate Poisson model (although only positive dependence can be captured by the bivariate Poisson distribution). Here, we follow [McHale and Scarf \(2011\)](#) who choose to allow for any potential dependence between the goals scored by the two teams by using a copula to ‘glue’ together the two marginal distributions of goals scored.

A copula,  $C$ , is a multivariate distribution with all univariate marginal distributions being uniformly distributed on the unit interval,  $[0, 1]$ . In other words,  $C$  is the distribution of a multivariate uniform random vector. The power of using a copula approach for modelling dependence comes from a theorem due to Sklar ([Sklar, 1973](#)) that states that the joint cumulative distribution function  $F$  of any pair of random variables  $(Y_1, Y_2)$  may be written in the form

$$F(y_1, y_2) = C(F_1(y_1), F_2(y_2)), \quad (y_1, y_2) \in \mathfrak{R}^2,$$

where  $F_1$  and  $F_2$  are their marginal cumulative distribution functions and  $C$  is a copula.

Here we want to join two marginal distributions and in this bivariate case, there are a plethora of copulas to choose from. We initially experimented with the Gaussian copula and Frank’s copula. However, Frank’s copula tended to provide a superior fit to the data according to the AIC and BIC. Of course, it also has the advantage of allowing for the full range of dependence so that the correlation can range from  $-1$  to  $1$ . The Frank copula is given by

$$C(u, v) = -\frac{1}{\kappa} \ln \left( 1 + \frac{(e^{-\kappa u} - 1)(e^{-\kappa v} - 1)}{e^{-\kappa} - 1} \right),$$

where  $\kappa \in \mathfrak{R}$  is the dependence parameter.

To construct our bivariate Weibull count model we use the Weibull count probability mass function given by (1) to first calculate the cumulative distribution functions,  $F_1(y_1; \lambda_1, c_1)$  and  $F_2(y_2; \lambda_2, c_2)$ . Using the copula  $C(u, v; \kappa)$  to glue these marginals together, the likelihood function for the parameter vector  $(\lambda_1, c_1, \lambda_2, c_2, \kappa)$  given the  $i$ th pair of observations  $(y_{1i}, y_{2i})$  is

$$\begin{aligned} \mathcal{L}(\lambda_1, c_1, \lambda_2, c_2, \kappa; y_{1i}, y_{2i}) &= \Pr(Y_1 = y_{1i}, Y_2 = y_{2i}) \\ &= C(F_1(y_{1i}), F_2(y_{2i})) - C(F_1(y_{1i} - 1), F_2(y_{2i})) - \\ &\quad C(F_1(y_{1i}), F_2(y_{2i} - 1)) + C(F_1(y_{1i} - 1), F_2(y_{2i} - 1)). \end{aligned} \tag{2}$$

The log-likelihood,  $\ell(\lambda_1, c_1, \lambda_2, c_2, \kappa; \mathbf{y}_1, \mathbf{y}_2) = \sum_{i=1}^n \log \mathcal{L}$ , for a sample of  $n$  football matches can be maximised using standard numerical optimization routines.

We note that Frank’s copula nests the independence case ( $\kappa = 0$ ) so that a test of whether  $\kappa$  is equal to 0 is equivalent to testing the assumption of independence.

### 2.3 A Model for Goals

Since [Maher \(1982\)](#) first gave his specification for modelling the goals scored by the two teams in a football match, many researchers have followed suit. Adapting the specification to our bivariate model

is simple due to the presence of the rate parameter,  $\lambda$ , in the distribution function given in equation (1). As in [Maher \(1982\)](#) we let the rate parameter for home team  $i$  playing against away team  $j$  be

$$\log(\lambda_i) = \alpha_i + \beta_j + \gamma,$$

where  $\alpha_i$  is the attack strength (the higher the value of  $\alpha$ , the stronger is the attack) of team  $i$ ,  $\beta_j$  is the defence strength (the smaller the value of  $\beta$ , the stronger is the defence) of team  $j$  and  $\gamma$  is a home advantage parameter. The away team's scoring rate is given by

$$\log(\lambda_j) = \alpha_j + \beta_i.$$

The above model is static in that the team strength parameters are not allowed to vary with time. This is not a good approximation. For example, [Baker and McHale \(2015\)](#) present a time varying model for team strengths and find that the strengths do indeed vary over time as, for example, teams buy and sell players, injuries occur and runs of good and bad form happen. In the forecasting literature there have been two approaches to allowing time-varying team strengths.

First, one can build a model in which the team strengths are assumed to vary stochastically, such as in [Crowder et al. \(2002\)](#), [Owen \(2011\)](#) or [Koopman and Lit \(2015\)](#). Second, one can adopt a time decay factor in the likelihood function so that more recent matches have greater weight in estimating the team strength parameters than matches further in the past. Implicitly one is assuming that the recent results of each team are indicative of that team's changing strength — a seemingly intuitive and reasonable assumption to make. This is the approach first presented in [Dixon and Coles \(1997\)](#) and is how we deal with the time-varying nature of team strengths here.

Using an exponential weighting function, our 'pseudo-likelihood',  $\tilde{\mathcal{L}}$ , at time  $t$  is given by

$$\tilde{\mathcal{L}}_t(\kappa, c, \alpha, \beta) = \prod_{k \in A_t} e^{-\xi(t-t_k)} \mathcal{L},$$

where  $t_k$  is the time when match  $k$  was played,  $A_t = \{k : t_k < t\}$  is the subset of all matches played up to, but not including, time  $t$  and  $\mathcal{L}$  is as given in equation (2). The parameter  $\xi$  cannot be estimated by maximising the likelihood. [Dixon and Coles \(1997\)](#) select the  $\xi$  that maximises

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

where  $\delta_k^H = 1$  if match  $k$  is a home win (else  $\delta_k^H = 0$ ), and  $p_k^H$ ,  $p_k^A$  and  $p_k^D$  are the maximum likelihood estimates of a home win, an away win and draw respectively. We modify this approach as our model is perhaps 'wasted' testing it against only the 1X2 (home win, draw, away win) market. Rather, we choose  $\xi$  to maximise the following objective function which includes the over-under 2.5 goals market

$$T(\xi) = \sum_{k=1}^N (\delta_k^H \log p_k^H + \delta_k^A \log p_k^A + \delta_k^D \log p_k^D + \gamma_k^{O2.5} \log p_k^{O2.5} + \gamma_k^{U2.5} \log p_k^{U2.5}), \quad (3)$$

where  $\gamma_k^{O2.5} = 1$  if there are more than 2.5 goals in match  $k$  and  $\gamma_k^{U2.5} = 1$  if there are fewer than 2.5 goals in match  $k$  and  $p_k^{O2.5}$  and  $p_k^{U2.5}$  are the model implied probabilities of there being more than or less than 2.5 goals in match  $k$ . [Figure 2](#) shows a plot of (3) as  $\xi$  varies from 0 to 0.008 and there is a clear maximum at  $\xi = 0.002$  and this is the value we use throughout. Following [Dixon and Coles \(1997\)](#)

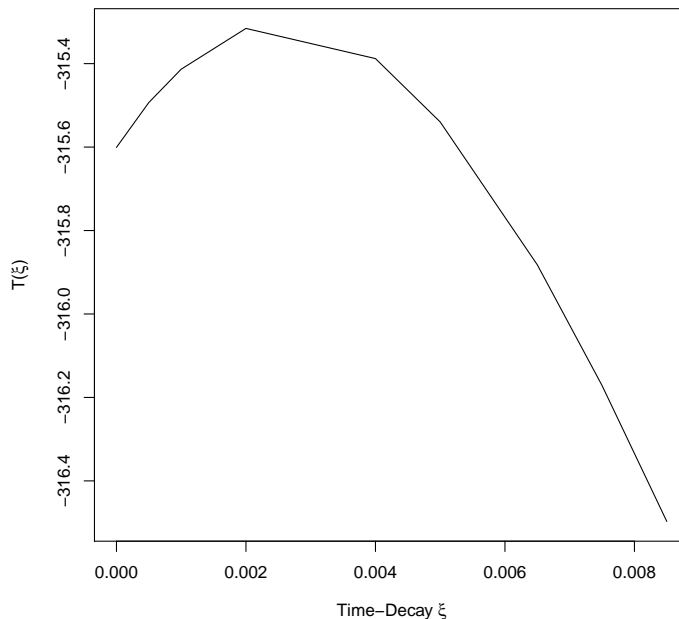


Figure 2: Objective function.

we use a half-week as the unit of time. They estimate a value of 0.0065 which is of the same order of magnitude of our estimated value of  $\xi$ .

### 3 Results

We obtained data on the results of matches in the English Premier League for the ten seasons from 2006/07 to 2015/16 inclusive from [www.football-data.co.uk](http://www.football-data.co.uk). The data also include pre-match odds from several bookmakers for the 1X2 (win, draw, lose) market and the over-under 2.5 goals market. We examine the goodness-of-fit of the models to these data using several methods: the log-likelihood and Akaike Information Criterion (AIC) are used for in-sample goodness-of-fit diagnostics, whilst returns to betting and calibration curves, are used to examine the out-of-sample predictive ability of the models. But first, we take a look at the estimated team strengths.

#### 3.1 Estimated Team Strengths

As we describe below, we fitted the model to rolling windows of four-and-a-half seasons. In this section we report the results at the end of the 2015/16 season. The estimated distribution parameters  $c$  in equation (1) are  $c_H = 1.050$  (0.0416) for the home team and  $c_A = 0.9831$  (0.0482) for the away team (standard errors in parentheses). The estimated value of the dependence parameter in the copula is  $\kappa = -0.4561$  (0.1961), so that the value of Kendall's  $\tau$  is  $-0.050$  suggesting a negative (and statistically significant) dependence. Lastly, the estimated home advantage is  $\gamma = 0.2948$  (0.0503). The standard errors have been computed assuming asymptotic normality and using numerical estimates of the gradient and the hessian. As pointed out by [McShane et al. \(2008\)](#) this gives reasonably close results compared to a bootstrap approach.

Table 2 shows the estimated team strength parameters,  $\alpha$  (attack) and  $\beta$  (defence), for our model. As one would expect, teams with fewer observations have larger standard errors for the estimated strengths. For example, Blackburn Rovers and Wolverhampton Wanderers have the two largest standard errors on their estimated attack strengths.

Table 2: Estimated team strength parameters, based on the last four-and-a-half seasons matches. Larger  $\alpha$ 's indicate stronger attack, smaller  $\beta$ 's stronger defence.

Team	$\alpha$ (s.e.)	$\beta$ (s.e.)
Arsenal	0.392 (0.076)	0.241 (0.103)
Aston Villa	-0.272 (0.101)	-0.281 (0.084)
Blackburn Rovers	-0.138 (0.316)	-0.422 (0.253)
Bolton Wanderers	-0.152 (0.310)	-0.379 (0.263)
Bournemouth	0.014 (0.156)	-0.283 (0.139)
Burnley	-0.470 (0.213)	-0.053 (0.165)
Cardiff City	-0.344 (0.221)	-0.423 (0.156)
Chelsea	0.371 (0.077)	0.196 (0.102)
Crystal Palace	-0.129 (0.107)	0.011 (0.105)
Everton	0.183 (0.083)	0.039 (0.095)
Fulham	-0.020 (0.129)	-0.373 (0.113)
Hull City	-0.261 (0.144)	-0.063 (0.127)
Leicester City	0.264 (0.104)	0.096 (0.125)
Liverpool	0.401 (0.076)	-0.013 (0.094)
Manchester City	0.550 (0.072)	0.242 (0.104)
Manchester United	0.312 (0.079)	0.230 (0.102)
Newcastle United	-0.058 (0.093)	-0.256 (0.085)
Norwich City	-0.201 (0.111)	-0.241 (0.095)
Queens Park Rangers	-0.173 (0.132)	-0.314 (0.110)
Reading	-0.103 (0.211)	-0.387 (0.172)
Southampton	0.165 (0.087)	0.120 (0.101)
Stoke City	-0.097 (0.093)	-0.013 (0.093)
Sunderland	-0.123 (0.094)	-0.128 (0.088)
Swansea City	0.024 (0.089)	-0.059 (0.091)
Tottenham Hotspur	0.302 (0.079)	0.106 (0.098)
Watford	-0.135 (0.167)	0.012 (0.156)
West Bromwich Albion	-0.111 (0.094)	-0.068 (0.091)
West Ham United	0.093 (0.089)	-0.019 (0.096)
Wigan Athletic	-0.001 (0.168)	-0.310 (0.153)
Wolverhampton Wanderers	-0.279 (0.338)	-0.682 (0.228)

### 3.2 In-sample Model Diagnostics

We compare the fit of our main model with the performance of three other models, an independent Poisson model, an independent Weibull count model and a Frank copula-induced bivariate Poisson model. Using these three models as benchmarks enables us to gauge where any out-performance may be originating (for example, an improvement in goodness-of-fit may come from modelling the dependence structure using a copula or it may come from modelling the counts using a Weibull count model rather than a Poisson distribution).

Table 3 shows the log-likelihood, the number of model parameters and the AIC for each of the four models under consideration. Although the copula-induced bivariate Weibull count model has more parameters, it is the best fitting model based on the AIC. It is noteworthy that the change from Poisson to

Weibull count distribution improves the AIC by approximately 7–8 units, and the change from independence to copula-induced dependence improves the AIC by approximately 12–14 units. As such, it looks like the overall improvement of our model comes from both the copula-based dependence and from the use of the Weibull count distribution.

Table 3: Comparison of the four models for football scores fitted (in-sample) to the Premier League data.

	Log-likelihood	Number of parameters	AIC
Copula Weibull Count Model	-3250.00	64	6628.00
Copula Poisson Model	-3257.09	62	6638.19
Independent Weibull Count Model	-3258.99	63	6643.98
Independent Poisson Model	-3264.00	61	6650.00

## 4 Out-of-sample Performance

To test the model out-of-sample, we fit the model to rolling windows of four-and-a-half seasons (1,710 matches), and for each fit predict the following week’s results. The first four-and-a-half season window begins at the start of the 2006/07 season and ends half way through the 2009/10 season. Having made predictions for the following week’s games, we move forward one week and refit the model to take account of the latest round of results. We repeat this until the last round of games in the 2009/10 season have been predicted. We then wait for half of the next season to be completed so that there are plenty of data from which reasonable parameter estimates can be obtained for the newly promoted teams and teams with large turnovers in playing staff. We repeat this process for each of the five seasons. This results in a total of 1,140 games (six  $\times$  half seasons) for which out-of-sample forecasts were generated and on which bets could be placed. Windows of length 4.5 seasons seem a good compromise between the desire to use more data for model fitting and keeping the model useful for prediction.

### 4.1 Calibration Curves for the Bivariate Weibull Count Model

Calibration can be intuitively seen as a way to visualise how often a model is right or wrong. In fact, a perfectly calibrated model *knows* how often it is right or wrong: when it predicts an event with 80% confidence, the event should occur 80% of the time. Whilst perfect accuracy for football forecasting models is probably an unachievable goal, perfect calibration is, in theory, a more realistic target, since a model that has imperfect accuracy could, in principle, be perfectly calibrated. Although popular in quantitative finance, the notion of calibration has never been investigated (to the best of our knowledge) in the sports forecasting literature.

In this section, we directly evaluate the calibration of the bivariate Weibull count model’s posterior prediction distribution using the 1,140 matches in our out-sample. For each event forecasted, we visualise the model’s performance graphically by plotting the *calibration curves* (also known as *reliability plots*). We now briefly describe how we estimate the calibration curves in football.

Consider a binary probabilistic prediction problem, which consists of binary labels and probabilistic predictions for them. Each instance has a *ground-truth label*  $y \in \{0, 1\}$  and an associated *predicted probability*  $q \in [0, 1]$  generated by the model, where  $q$  represents the model’s posterior probability of the instance having a positive label ( $y = 1$ ). The calibration curve is simply a plot of the label frequency,  $P(y = 1|q)$ , versus predicted probability. However, computing  $P(y = 1|q)$  requires an infinite amount of



data and hence approximation methods are needed to perform the calibration analysis. We follow here Tukey's (Tukey et al., 1961) approach and divide the prediction space by 'halves': we split the data into upper and lower halves, then split those halves, then split the extreme halves recursively. Compared to equal-width binning, this allows visual inspection of tail behaviour without devoting too many graphical elements to the bulk of the data. A perfectly calibrated curve would coincide with the  $y = x$  line, so that the empirical frequency of an event equalled the model estimated probability. When the curve lies above the diagonal, the model is *pessimistic* in that it under-estimates the probability of the event occurring; and when it is below the diagonal, the model is *optimistic* in that it over-estimates the probability of the event occurring.

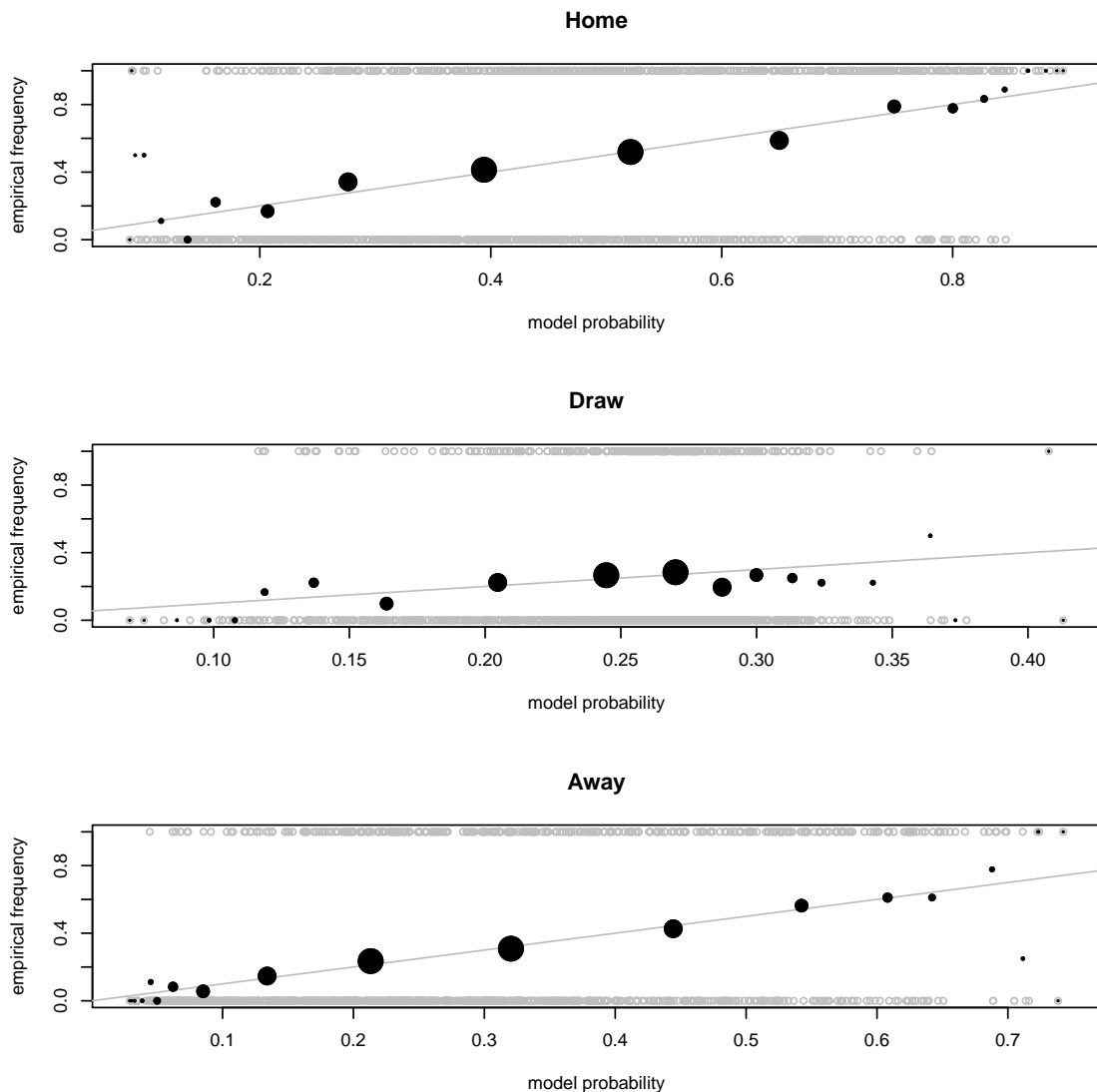


Figure 3: Calibration curves for predicting outcomes in the 1X2 market. The size of the circles are proportional to the number of observations in each bin.

The calibration curves for predicting home win, draw and away win outcomes in the 1X2 market are shown in Figure 3. Overall it appears that the model is 'well-calibrated' - the points lie near the  $y = x$  line.

## 4.2 Betting Performance

We now test all four models against the betting market. There is a vast array of work in the economics literature examining the efficiency of the betting market on football, and, on the whole, there is agreement that the market is efficient in that it is not possible to accrue ‘superior’ returns (see, for example, (Snowberg and Wolfers, 2010)). Thus, comparing the probabilities implied in the betting market with those produced by the model is a simple, but informative guide to the model’s effectiveness.

Our betting simulation is *out-of-sample*: team strengths are estimated using results prior to the match to be bet on. As a consequence of the efficient markets hypothesis, we would consider a return of near the market over-round as evidence that a model is working well. We use the average odds available on two markets: the 1X2 (home win, draw, away win) market, and the over-under 2.5 goals market. During the last half of each of the ten seasons of data, the average over-rounds on the two markets were 5.5% and 6.0% respectively. By testing our model against the over-under market we are gaining an understanding of the model’s performance in predicting what it was designed to forecast - goals. If we were to test the model against only the 1X2 market, we would be disregarding the main output from the model - the probabilities of each and every possible scoreline.

Our investment strategy is based on the Kelly Criterion (Kelly, 1956). The Kelly Criterion is borne from a desire to maximise long-run log-utility and it results in an investment strategy where the bettor invests a fraction  $f$  of his overall wealth

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

where  $p$  is the bettor’s estimate of the probability of an event (e.g. the home team winning the game), and  $b$  is the (fractional) odds offered by the bookmaker (where  $1/(b + 1)$  can be interpreted loosely as the bookmaker’s implied probability of the event occurring).

We allow a maximum of 1 units per bet and use the Kelly criterion to decide on what fraction of our 1 unit is staked. Effectively we reset our bankroll to 1 after each bet. An additional ‘protection’ was also introduced: we restrict ourselves to ‘quality bets’ when the expected value of any bet is greater than some threshold. For each game, there are five possible events to bet on: home win, draw, away win, over 2.5 goals and under 2.5 goals. For event type  $A$ , we only bet if

$$EV(A) = P(A) \times Odds(A) - 1 > t,$$

where  $t$  is the threshold parameter. Experimentation led us to use  $t = 0.04$  which is a good compromise between betting too much (and losing) and placing a reasonable number of bets. Table 4 shows the *out-of-sample* returns to 1X2 betting on matches in the out-sample of 1,140 matches and Table 5 shows the *out-of-sample* returns to betting on the over-under 2.5 goals market for the same 1,140 games.

Table 4: Summary of results when betting on the 1X2 market using a Kelly betting strategy.

	Model	Number of bets	Number of winning bets	Gross return	Net return	Total staked	Total return
	Copula Weibull Count	662	303	675.01	13.01	64.05	20.3%
	Independent Weibull Count	707	294	716.11	9.11	66.53	13.7%
	Copula Poisson Model	683	289	690.42	7.42	58.83	12.6%
	Independent Poisson Model	740	288	747.66	7.66	62.78	12.2%

In both cases the copula Weibull count model produces the highest returns: 20.3% in the case of the

Table 5: Summary of results when betting on the over-under 2.5 goals market using a Kelly betting strategy.

	Model	Number of bets	Number of winning bets	Gross return	Net return	Total staked	Total return
	Copula Weibull Count	403	214	409.11	6.11	40.68	15.0%
	Independent Weibull Count	429	224	432.79	3.79	45.28	8.4%
	Copula Poisson Model	411	211	415.11	4.11	42.28	9.7%
	Independent Poisson Model	418	217	422.26	4.26	44.25	9.6%

1X2 market and 15.0% in the over-under market. Reading from the bottom row of Table 4 to the top row shows how improvements in the returns are gained from first adopting the copula bivariate model over the independent model and second by using the Weibull count distribution instead of the Poisson distribution.

The profits demonstrated above may be obtained as a consequence of using an improved model, or an improved betting strategy. To illuminate this issue, we also tested the models using a simple one unit betting strategy - if the expected value of the bet was above the threshold  $t$ , we bet one unit. Note that this strategy will stake the same amount independently of the value of the expected value (as long as it exceeds the threshold). Doing so resulted in returns from the copula Weibull count model of 22.9% - higher than for the Kelly betting strategy. This suggests that the returns are coming from the model, and not necessarily from the betting strategy, although we note that we still prefer the use of the Kelly betting strategy because it has smaller confidence intervals and a higher realised Sharpe ratio (0.096 for the Kelly strategy versus 0.050 for the unit betting strategy) which makes its returns more ‘robust’.

Before moving away from testing our model against the betting market, we present the evolution of the cumulative returns over the betting period, together with a bootstrapped confidence interval. Using a bootstrap confidence interval will allow us to see just how ‘robust’ the achieved returns are. To compute the confidence intervals we utilised a bootstrap procedure in which 100 replicate datasets of identical size to the original (1,140 games) were generated by sampling individual predicted probabilities-odds pairs,  $(p_i, O_i)$ , and the eventual outcome, with replacement. The results reported for the (95%) confidence intervals are based on a normal approximation and standard errors estimated by the standard deviation of the returns across those samples.

Figure 4 shows the cumulative profit according to the Kelly betting strategy on the 1X2 market. The plot is similar for the over-under 2.5 goals market. The flat regions are when we do not bet. The most promising sign here is that the lower bound of the confidence interval is positive.

Obtaining returns that are superior and positive in both the 1X2 and over-under 2.5 markets, is interesting for two reasons. First, we take the results as validation of our model, and second, and more profoundly, there are implications for market efficiency, which may be the subject of future research.

## 5 Discussion

In this paper we have studied a new model for bivariate counts to predict the score distribution in football matches. Our model assumes that the distribution of inter-arrival time for goals is Weibull, rather than exponential (the latter is implied when using the Poisson distribution). We induce dependency between the two marginals using a Frank copula. We test this bivariate Weibull count model against three alternative models both in-sample and out-of-sample and the bivariate Weibull count model provides a superior fit to results data from the English Premier League.

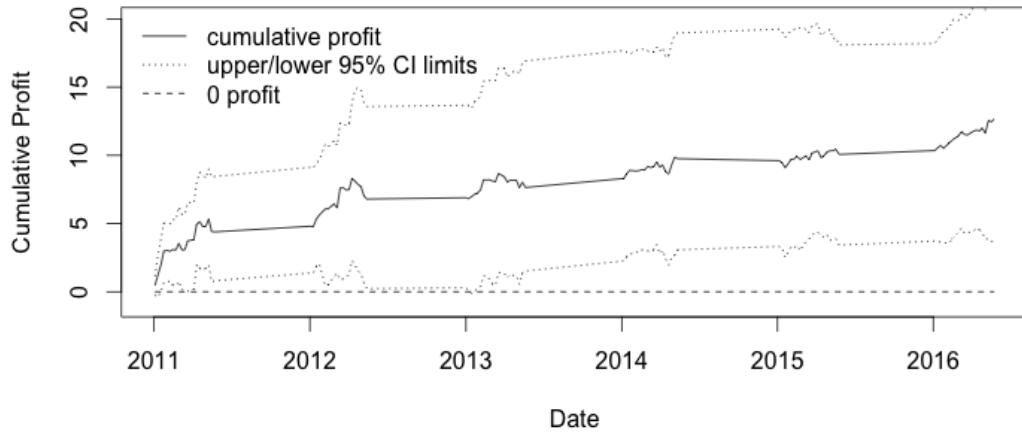


Figure 4: Cumulative profit to Kelly betting strategy on the 1X2 market using the bivariate Weibull count model.

Perhaps most interesting is the finding that our model attains positive returns in both the 1X2 and over-under 2.5 goals betting markets. However, it is worth noting that although we can ‘beat the bookmakers’, we are picking and choosing a relatively small number of matches to bet on. A bookmaker must produce competitive odds for an entire league fixture list. To do so is a more challenging task and requires intimate knowledge of the leagues, teams and players involved. Hence, in the betting industry, bookmakers remain reliant on a system in which both statistical models and traders work in tandem to produce odds. However, the model presented here may prove to be of use to bookmakers. For example, we believe that bookmakers estimate a supremacy rating for one team over another in a match and an expected number of goals in the match. Bookmakers could use these two variables, which are based on intimate knowledge of the form of the teams and the players actually playing for each team, as inputs to our model. The 1X2 markets will still be priced the same but by better estimating the goals distribution, more accurate prices for the subsidiary markets will be produced.

In future work we hope to bridge the gap between the statistical model and traders by incorporating player skills into our model. Using knowledge on the identity of the players on each team (and their abilities) should avoid the need for estimating time-varying team strengths as the key component of the dynamic nature of team strengths is the changing line-up of the team. If accurate player ratings are used as inputs to the model, such an approach should produce improved forecasts as it will better capture the expected scoring intensities of each team.

## Acknowledgements

We would like to thank the two anonymous reviewers and the Associate Editor for their comments and suggestions in helping to improve the paper.

## References

- Baker, R. D. and McHale, I. G. (2015). Time varying ratings in association football: the all-time greatest team is.. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 178(2):481–492.
- Crowder, M., Dixon, M., Ledford, A., and Robinson, M. (2002). Dynamic modelling and prediction of english football league matches for betting. *Journal of the Royal Statistical Society: Series D (The Statistician)*, 51(2):157–168.
- Dixon, M. J. and Coles, S. G. (1997). Modelling association football scores and inefficiencies in the football betting market. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 46(2):265–280.
- Eddelbuettel, D. and Sanderson, C. (2014). Rcpparmadillo: Accelerating r with high-performance c++ linear algebra. *Computational Statistics and Data Analysis*, 71:1054–1063.
- Karlis, D. and Ntzoufras, I. (2003). Analysis of sports data by using bivariate poisson models. *Journal of the Royal Statistical Society: Series D (The statistician)*, 52:381–393.
- Kelly, J. L. (1956). A new interpretation of information rate. *Bell System Technical Journal*, 35(4):917–926.
- Koopman, S. J. and 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.
- Maher, M. J. (1982). Modelling association football scores. *Statistica Neerlandica*, 36(3):109–118.
- McHale, I. G. and Scarf, P. (2011). Modelling the dependence of goals scored by opposing teams in international soccer matches. *Statistical Modelling*, 11(3):219–236.
- McShane, B., Adrian, M., Bradlow, E. T., and Fader, P. S. (2008). Count models based on weibull interarrival times. *Journal of Business and Economic Statistics*, 26(3):369–378.
- Owen, A. (2011). Dynamic bayesian forecasting models of football match outcomes with estimation of the evolution variance parameter. *IMA Journal of Management Mathematics*, 22:99–113.
- R Core Team (2014). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.
- Rue, H. and Salvesen, O. (2000). Prediction and retrospective analysis of soccer matches in a league. *Journal of the Royal Statistical Society: Series D (The statistician)*, 49:399–418.
- Sklar, A. (1973). Random variables, joint distribution functions, and copulas. *Kybernetika*, 9(6):449–460.
- Snowberg, E. and Wolfers, J. (2010). Explaining the favorite-longshot bias: is it risk-love or misperceptions? *Journal of Political Economy*, 118:723–746.
- Tukey, J. W. et al. (1961). Curves as parameters, and touch estimation. In *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability, Volume 1: Contributions to the Theory of Statistics*. The Regents of the University of California.