Journal of the Royal Statistical Society
SERIES A Statistics in Society

J. R. Statist. Soc. A (2014)

A dynamic bivariate Poisson model for analysing and forecasting match results in the English Premier League

Siem Jan Koopman and Rutger Lit

VU University Amsterdam, The Netherlands

[Received September 2012. Revised July 2013]

Summary. We develop a statistical model for the analysis and forecasting of football match results which assumes a bivariate Poisson distribution with intensity coefficients that change stochastically over time. The dynamic model is a novelty in the statistical time series analysis of match results in team sports. Our treatment is based on state space and importance sampling methods which are computationally efficient. The out-of-sample performance of our methodology is verified in a betting strategy that is applied to the match outcomes from the 2010–2011 and 2011–2012 seasons of the English football Premier League. We show that our statistical modelling framework can produce a significant positive return over the bookmaker's odds.

Keywords: Betting; Importance sampling; Kalman filter smoother; Non-Gaussian multivariate time series models; Sport statistics

1. Introduction

The prediction of a football match is a challenging task. The pundit usually has strong beliefs about the outcomes of games. Bets can be placed on a win, a loss, a draw or on the match score itself. The collection of the predictions is reflected by the bookmaker's odds. In this paper we study the history of 9 years of football match results from the English Premier League. The number of goals scored by a team may depend on the strength of attack of the team, the strength of defence of the opposing team, the home ground advantage (when applicable) and the development of the match itself. We analyse the match results on the basis of a dynamic statistical modelling framework in which the strengths of attack and defence of the teams can vary over time. We show that the forecasts from this model are sufficiently accurate to gain a positive return over the bookmaker's odds.

Many statistical analyses of match results are based on the product of two independent Poisson distributions, which is also known as the double-Poisson distribution. The means of the two distributions can be interpreted as the goal scoring intensities of the two competing teams. In our modelling framework, the bivariate Poisson distribution is used which includes a dependence parameter that allows for correlation between home and away match scores. It represents the phenomenon that the ability or the effort of a team for a particular game is influenced by the other team or by the way that the match progresses. The performances of the teams due to the interactions between teams are captured by the dependence parameter. Furthermore, we let the goal scoring intensities of the two teams depend on the strengths of attack and defence of the

Address for correspondence: Siem Jan Koopman, Department of Econometrics, VU University Amsterdam, De Boelelaan 1105, 1081 HV Amsterdam, The Netherlands.
E-mail: s.j.koopman@vu.nl

0964–1998/14/177000

two teams. These strengths for each team are allowed to change stochastically over time. This time varying feature becomes more important when we jointly analyse the match results for a series of consecutive football seasons. For example, when an excellent scorer leaves a team, the strength of attack weakens. Overall we expect that strengths of attack and defence change slowly over time.

The basis of our modelling approach was proposed by Maher (1982). In this study, the double-Poisson distribution, with the means expressed as team-specific strengths of attack and defence, is adopted as the underlying distribution for goal scoring. Maher (1982) explored the existence of a small correlation between home and away match scores; he found a considerable improvement in model fit by trying a range of values for the dependence parameter. He did not provide parameter estimates of the correlation or dependence parameter. Furthermore, Maher's basic model is static; the team's strengths of attack and defence do not vary over time. Dixon and Coles (1997) considered the double-Poisson model with a dependence parameter that is estimated together with the other parameters. They suggested that the assumption of independence between goal scoring is reasonable except for the match results 0–0, 1–0, 0–1 and 1–1. They also introduced a weighting function to downweight likelihood contributions of observations from the more distant past. Karlis and Ntzoufras (2003) also used a bivariate Poisson distribution for match results; they showed that even a small value for the dependence parameter leads to a more accurate prediction of the number of draws. However, attack and defence strengths are kept static over time in their analysis. Rue and Salvesen (2000) incorporated the framework of Dixon and Coles (1997) within a dynamic generalized linear model and adopted Markov chain Monte Carlo methods to study the time varying properties of the football teams in continuous time. In their analysis of match results, they truncated the number of goals to a maximum of 5 because they argued that the number of goals beyond 5 provides no further information about the strengths of attack and defence of a team. Owen (2011) adopted a similar dynamic generalized linear model and also used Markov chain Monte Carlo methods for estimation. However, he argued strongly for a model in discrete time. He found that the evolution of parameters over time, the role of attack and defence strengths and the effect of home and away match scores are more effectively analysed in discrete time. We also formulate the model in discrete time but our model is based on the bivariate Poisson distribution and we estimate the parameters by simulated maximum likelihood methods rather than Markov chain Monte Carlo methods. Owen (2011) empirically verified the model for a low dimensional data set whereas we consider all matches in the English Premier League for 9 years.

The following contributions in the literature involve multivariate time series models and sports but are less relevant to our study. Ord *et al.* (1993) considered a moderate multivariate extension of a Bayesian dynamic count data model for the analysis and forecasting of the number of goals scored by a small number of teams over a period of time. Furthermore their modelling framework is not based on Maher (1982). In Crowder *et al.* (2002), the dynamic generalized linear model of Dixon and Coles (1997) is formulated as a non-Gaussian state space model with time varying strengths of attack and defence as well. However, they used approximate methods for parameter estimation as they stated that an exact analysis is computationally too expensive. Given the rapid development of simulation methods for time series models, we shall show that exact maximum likelihood methods for an extensive analysis of match results can be carried out as a matter of routine. Our empirical study aims to analyse match results from the English Premier League. Earlier and leading studies have analysed match results from other sport leagues. In particular, Glickman and Stern (1998) and Glickman (2001) have considered match results from the American Football League, Fahrmeir and Tutz (1994) from the German *Bundesliga* and Knorr-Held (2000) from the American National Basketball Association.

We show that football match results from a high dimensional data set can be analysed effectively within a non-Gaussian state space model where the observed pairs of counts are assumed to come from a bivariate Poisson distribution. We have attack and defence strengths that are stochastically evolving over time. The statistical analysis is based on exact maximum likelihood and signal extraction methods which rely on efficient Monte Carlo simulation techniques such as importance sampling. Several extensions can be considered within our modelling framework. For example, we introduce a parameter that accounts for the transition of summer and winter breaks. We also introduce the diagonal inflation method of Dixon and Coles (1997) for the bivariate Poisson distribution to account for the overrepresentation of draws. Finally we emphasize that we do not need to truncate the observed match outcomes to some maximum value in our analysis.

The remainder of the paper is organized as follows. Our dynamic statistical modelling framework for the bivariate Poisson distribution is introduced and discussed in detail in Section 2. It is shown how we can represent the dynamic model in a non-Gaussian state space form. The statistical analysis relies on advanced simulation-based time series methods which are developed elsewhere. We provide the implementation details and some necessary modifications of the methods. The analysis includes maximum likelihood estimation, signal extraction of the strengths of attack and defence of a team and the forecasting of match results. In Section 3 we illustrate the methodology for a high dimensional data set of football match results from the English Premier League during the seasons from 2003–2004 to 2011–2012. The first seven seasons are used for parameter estimation and in-sample diagnostic checking of the empirical results whereas the last two seasons are used for the out-of-sample forecast evaluation of the model. A forecasting study is presented in Section 4 where we give evidence that our model can turn a positive return over the bookmakers' odds by applying a simple betting strategy during the seasons of 2010–2011 and 2011–2012. Concluding remarks are given in Section 5.

The data that are analysed in the paper and a selection of the programs that were used to analyse them can be obtained from

`http://wileyonlinelibrary.com/journal/rss-datasets`

2. Statistical modelling framework

We analyse football match results in a competition for a number of seasons as a time series panel of pairs of counts. We assume that an even number of J teams play in a competition and hence each week $J/2$ matches are played. It also follows that a season consists of $2(J-1)$ weeks in which each team plays against another team twice, as a home team and as a visiting team. The specific details of our data set for the empirical study are discussed in Section 3.

2.1. Bivariate Poisson model

The result or outcome of a match between the home football team i and the visiting football team j in week t is taken as the pair of counts $(X, Y) = (X_{it}, Y_{jt})$, for $i \neq j = 1, \ldots, J$ and $t = 1, \ldots, n$, where n is the number of weeks available in our data set. The first count X_{it} is the non-negative number of goals scored by the home team i and the second count Y_{jt} is the number of goals scored by the visiting team j, in week t. Each pair of counts (X, Y) is assumed to be generated or sampled from the bivariate Poisson distribution with probability density function

$$p_{\mathrm{BP}}(X, Y; \lambda_x, \lambda_y, \gamma) = \exp(-\lambda_x - \lambda_y - \gamma) \frac{\lambda_x^X}{X!} \frac{\lambda_y^Y}{Y!} \sum_{k=0}^{\min(X,Y)} \binom{X}{k} \binom{Y}{k} k! \left(\frac{\gamma}{\lambda_x \lambda_y} \right)^k, \qquad (1)$$

for $X = X_{it}$ and $Y = Y_{jt}$, with λ_x and λ_y being the intensities for X and Y respectively, and γ being a coefficient for the dependence between the two counts in the pair, X and Y. In short notation, we write

$$(X, Y) \sim \mathrm{BP}(\lambda_x, \lambda_y, \gamma).$$

The means, variances and covariance for the home team score X and the away team score Y are

$$\left.\begin{aligned} \mathbb{E}(X) = \mathrm{var}(X) = \lambda_x + \gamma, \\ \mathbb{E}(Y) = \mathrm{var}(Y) = \lambda_y + \gamma, \\ \mathrm{cov}(X, Y) = \gamma, \end{aligned}\right\} \tag{2}$$

and hence the correlation coefficient between X and Y is given by

$$\rho = \frac{\gamma}{\sqrt{\{(\lambda_x + \gamma)(\lambda_y + \gamma)\}}}.$$

This definition of the bivariate Poisson distribution is not unique: other formulations have also been considered; see, for example, Kocherlakota and Kocherlakota (1992) and Johnson *et al.* (1997). A different formulation for the bivariate Poisson distribution also implies different means, variances and covariances in expression (2).

The difference between the counts X and Y determines whether the match is a win, a loss or a draw for the home team. The variable $X - Y$ has a discrete probability distribution known as the Skellam distribution and it is invariant to γ when $(X, Y) \sim \mathrm{BP}(\lambda_x, \lambda_y, \gamma)$ for $\gamma > 0$; see Skellam (1946). Karlis and Ntzoufras (2006, 2009) have used the Skellam distribution to analyse differences in scores in soccer matches.

2.2. Dynamic specification for goal scoring intensities

The scoring intensities of two teams playing against each other are determined by λ_x, λ_y and γ. In our modelling framework, we let λ_x and λ_y vary with the pairs of teams that play against each other. Furthermore, we allow these intensities to change slowly over time since the composition and the performance of the teams will change over time. The intensity of scoring for team i, when playing against team j, is assumed to depend on the strength of attack of team i and the strength of defence of team j. We also acknowledge the home ground advantage in scoring by having the coefficient δ; this relative advantage is considered to be the same for all teams and constant over time. In Section 2.3, we introduce a model extension in which δ is not the same for all teams. The strength of attack of the home team i in week t is denoted by α_{it} and its strength of defence is denoted by β_{it} for $i = 1, \ldots, J$. The goal scoring intensities for home team i and away team j in week t are then specified as

$$\begin{aligned} \lambda_{x,ijt} &= \exp(\delta + \alpha_{it} - \beta_{jt}), \\ \lambda_{y,ijt} &= \exp(\alpha_{jt} - \beta_{it}). \end{aligned} \tag{3}$$

In a football season with $J(J-1)$ matches, $2J(J-1)$ goal counts and for some time index t, we can identify the unknown signals for attack α_{it}s and defence β_{it}s together with coefficient δ, i.e. $2J + 1$ unknowns, when the number of teams is $J > 2$. The time variation of the strengths of attack and defence can be identified when we analyse match results from a series of football seasons.

All teams in the competition are assumed to have unique strengths of attack and defence which we do not relate to each other. In effect we assume that each team can compose their teams independently of each other. The strengths of attack and defence of the team can change over time since the composition of the team will not be constant over time. Also the performances of the teams are expected to change over time. We therefore specify the strengths of attack and defence as auto-regressive processes. We have

$$\begin{aligned}\alpha_{it} &= \mu_{\alpha,i} + \phi_{\alpha,i}\alpha_{i,t-1} + \eta_{\alpha,it},\\ \beta_{it} &= \mu_{\beta,i} + \phi_{\beta,i}\beta_{i,t-1} + \eta_{\beta,it},\end{aligned} \tag{4}$$

where $\mu_{\alpha,i}$ and $\mu_{\beta,i}$ are unknown constants, $\phi_{\alpha,i}$ and $\phi_{\beta,i}$ are auto-regressive coefficients and the disturbances $\eta_{\alpha,it}$ and $\eta_{\beta,it}$ are normally distributed error terms which are independent of each other for all $i = 1, \ldots, J$ and all $t = 1, \ldots, n$. We assume that the dynamic processes are independent of each other and that they are stationary. It requires that $|\phi_{\kappa,i}| < 1$ for $\kappa = \alpha, \beta$ and $i = 1, \ldots, J$. The independent disturbance sequences are stochastically generated by

$$\eta_{\kappa,it} \sim \mathrm{NID}(0, \sigma^2_{\kappa,i}), \qquad \kappa = \alpha, \beta, \tag{5}$$

where $\mathrm{NID}(c, d)$ refers to normally independently distributed with mean c and variance d, for $i = 1, \ldots, J$ and $t = 1, \ldots, n$.

The initial conditions for the auto-regressive processes α_{it} and β_{it} can be based on means and variances of their unconditional distributions, which are given by

$$\mathbb{E}(\kappa_{it}) = \mu_{\kappa,i}/(1 - \phi_{\kappa,i}), \qquad \mathrm{var}(\kappa_{it}) = \sigma^2_{\kappa,i}/(1 - \phi^2_{\kappa,i}), \qquad \kappa = \alpha, \beta.$$

Other, and possibly more complicated, dynamic structures for α_{it} and β_{it} can be considered as well but in our current study we shall consider only the first-order auto-regressive processes as given in expression (4).

2.3. Some extensions of the basic model

Our basic modelling framework for match results can be extended in several directions. First, we address the tendency of the bivariate Poisson distribution (1) to underestimate draws in match results, in particular when $\gamma = 0$, i.e. the double-Poisson model. For example, Dixon and Coles (1997) found that the scores 0–0 and 1–1 were underrepresented in their extended data set in favour of 1–0 and 0–1. They proposed to adjust their double-Poisson model by introducing an adjustment term that shifts probability mass from 1–0 and 0–1 towards 0–0 and 1–1. The adjustment is referred to as diagonal inflation and we apply it to the bivariate Poisson density function (1). The resulting density function is obtained by multiplying the term $\pi_{\lambda_x,\lambda_y}(X, Y)$ with density function (1) where

$$\pi_{\lambda_x,\lambda_y}(X, Y) = \begin{cases} 1 + \lambda_x\lambda_y\omega, & \text{if } (X, Y) = (0, 0),\\ 1 - \lambda_x\omega, & \text{if } (X, Y) = (0, 1),\\ 1 - \lambda_y\omega, & \text{if } (X, Y) = (1, 0),\\ 1 + \dfrac{\omega}{1 + \gamma/\lambda_x\lambda_y}, & \text{if } (X, Y) = (1, 1),\\ 1, & \text{otherwise,}\end{cases} \tag{6}$$

where coefficient ω determines how much probability mass is shifted. The multiplication leads to a proper density with moments that are the same as those of the bivariate Poisson distribution. A different but related adjustment was considered by Karlis and Ntzoufras (2003).

Another extension of our basic model is to allow for summer and winter breaks in league matches. In most football leagues, players can be bought or hired only during the summer and winter breaks. A change in the composition of a team can lead to changes in their strengths of attack and defence. We allow for such changes in the paths of α_{it} and β_{it} by letting the random shocks $\eta_{\alpha,it}$ and $\eta_{\beta,it}$ respectively have different scalings for the first time period after the summer and winter breaks. When the processes for α_{it} and β_{it} are sufficiently persistent, large random shocks will lead to breaks in these processes. Hence we replace the distributions for the disturbance sequences $\eta_{\kappa,it}$ in expression (5) by

$$\eta_{\kappa,it} \sim \text{NID}\{0, \sigma^2_{\kappa,i} + \sigma^2_{\kappa,\text{S}}\tau_{\text{S}}(t) + \sigma^2_{\kappa,\text{W}}\tau_{W}(t)\}, \qquad \kappa = \alpha, \beta, \tag{7}$$

for team $i = 1, \ldots, J$, where the indicator variables $\tau_{\text{S}}(t)$ and $\tau_{\text{W}}(t)$ are set equal to 1 at the end of the summer and winter breaks respectively, and to 0 otherwise, with $\sigma^2_{\kappa,\text{S}} > 0$ and $\sigma^2_{\kappa,\text{W}} > 0$. As a result, all disturbance variances are time varying. The two additional variances for the strengths of both attack and defence are estimated jointly with the other parameters in ψ of equation (18) in Section 3.2.

Our final extension concerns the home ground advantage δ which is the same for all teams. It is realistic to expect that the home ground advantage has different effects on different teams. By introducing a team-specific home ground advantage in the model, the number of parameters increases and it will slow down the estimation process. A more feasible option is to limit this extension by pooling home ground advantage coefficients for groups of teams. For example, in Section 5 we consider a different home ground coefficient for the traditionally well performing teams in the English Premier League: Arsenal, Chelsea, Liverpool, Manchester City and Manchester United. We may expect that for these teams the effect of home ground advantage on match results is higher. An interesting discussion of what home ground advantage represents is given by Pollard (2008).

2.4. *General state space representation of the model*

For our model-based analysis, it is convenient to present the model in the general state space form. The pair (X_{it}, Y_{jt}) is the observed outcome of the match of home team i against the visiting team j which is played at time t. The statistical dynamic model for the match result (X_{it}, Y_{jt}) of home team i against team j is given by

$$(X_{it}, Y_{jt}) \sim \text{BP}(\lambda_{x,ijt}, \lambda_{y,ijt}, \gamma), \tag{8}$$

where BP refers to the bivariate Poisson distribution with density function (1) and with the goal scoring intensities $\lambda_{x,ijt}$ and $\lambda_{y,ijt}$ specified via the link functions

$$\lambda_{x,ijt} = s_{x,ij}(z_t), \qquad \lambda_{y,ijt} = s_{y,ij}(z_t), \qquad i \neq j = 1, \ldots, J.$$

Here the so-called state vector z_t contains the strengths of attack and defence of all J teams at time t, i.e.

$$z_t = (\alpha_{1t}, \ldots, \alpha_{Jt}, \beta_{1t}, \ldots, \beta_{Jt})', \qquad t = 1, \ldots, n. \tag{9}$$

Hence the dimension of the state vector is $2J \times 1$. We can represent the goal scoring intensity specifications in expression (3) by having the link functions as

$$s_{x,ij}(z_t) = \exp(\delta + w_{ij} z_t), \qquad s_{y,ij}(z_t) = \exp(w_{ji} z_t), \qquad i \neq j = 1, \ldots, J, \tag{10}$$

where w_{ij} selects the appropriate α_{it} and β_{jt} elements from z_t in expression (9). The transformation of the state vector into goal scoring intensities is illustrated in the on-line appendix. The bivariate Poisson distribution that is used in expression (8) relies further on dependence coefficient γ and $s_{x,ij}(z_t)$ relies also on the home ground advantage coefficient δ. We collect such unknown coefficients in the parameter vector ψ for which more details are given below.

The linear dynamic process for the $2J \times 1$ state vector is given generally by

$$z_t = \mu + \Phi z_{t-1} + \eta_t, \qquad \eta_t \sim \text{NID}(0, H), \tag{11}$$

for $t = 1, \ldots, n$, where μ is the constant vector of dimension $2J \times 1$, Φ is the auto-regressive coefficient matrix of dimension $2J \times 2J$ and the disturbance vector η_t of dimension $2J \times 1$ is normally independently distributed with mean 0 and variance matrix H. The vector μ and matrices Φ and H may rely partly on unknown coefficients which we also collect in the parameter vector ψ. The initial condition for the state vector z_1 can be obtained from the unconditional properties of z_t.

The dynamic specifications of the strengths of attack and defence in expression (4) can be represented in the general form of expression (11) as follows. We collect the disturbances of expression (4) in $\eta_t = (\eta_{\alpha,1t}, \ldots, \eta_{\alpha,Jt}, \eta_{\beta,1t}, \ldots, \eta_{\beta,Jt})'$. Next we need to define only the matrices μ, Φ and H as

$$\mu = (\mu_{\alpha,1}, \ldots, \mu_{\alpha,J}, \mu_{\beta,1}, \ldots, \mu_{\beta,J})',$$
$$\Phi = \mathrm{diag}(\phi_{\alpha,1}, \ldots, \phi_{\alpha,J}, \phi_{\beta,1}, \ldots, \phi_{\beta,J}),$$
$$H = \mathrm{diag}(\sigma^2_{\alpha,1}, \ldots, \sigma^2_{\alpha,J}, \sigma^2_{\beta,1}, \ldots, \sigma^2_{\beta,J}),$$

where $\mathrm{diag}(v)$ refers to a diagonal matrix with the elements of v on the leading diagonal. The constant vector μ is captured in the unknown initial state vector z_1. The remaining unknown coefficients are then placed in the parameter vector ψ. In this case we have

$$\psi = (\phi', h', \delta, \gamma)',$$

where the column vectors ϕ and h contain the diagonal elements of Φ and H respectively. It can imply that the number of unknown coefficients is large and the burden of parameter estimation is high. In practice, we shall pool many unknown coefficients into a smaller set of parameters. This is illustrated in our empirical study of Section 3.

2.5. *Evaluation of likelihood function and estimation*

We opt for the method of maximum likelihood to obtain parameter estimates with optimal properties in large samples. Hence we require to develop an expression for the likelihood function of our model. For the evaluation of the likelihood function we require simulation methods because the multivariate model is non-Gaussian and non-linear and hence we cannot rely on linear estimation methods for dynamic models such as the Kalman filter.

We have $J/2$ match results for each week t. A specific match result is denoted by (X_{it}, Y_{jt}) with $i \neq j$ and $i, j \in \{1, \ldots, J\}$. The numbers of goals scored by all teams in week t are collected in the $J \times 1$ observation vector y_t. The observation density of y_t for a given realization of the state vector z_t is then given by

$$p(y_t | z_t; \psi) = \prod_{k=1}^{J/2} p_{\mathrm{BP}}(\lambda_{x,ijt}, \lambda_{y,ijt}, \gamma), \tag{12}$$

where p_{BP} is the probability density function (1) of the bivariate Poisson distribution and where index k represents the kth match between home team i against visiting team j. We note that $\lambda_{x,ijt} = s_{x,ij}(z_t)$ and $\lambda_{y,ijt} = s_{y,ij}(z_t)$ where the link functions can, for example, be based on expression (3). In this case we can express the signal vector that is associated with the density $p(y_t | z_t; \psi)$ as

$$\mathbb{E}(y_t | z_t; \psi) = \exp(a_t \delta + W_t z_t), \tag{13}$$

where vector a_t consists of elements equal to 1 when the scores of the corresponding elements in y_t are from the home team and 0 otherwise, whereas matrix W_t is composed of the appropriate row vectors w_{ij} as introduced in expression (10). The home ground advantage coefficient δ is part of the parameter vector ψ.

We define $y=(y_1',\ldots,y_n')'$ and $z=(z_1',\ldots,z_n')'$ for which it follows that

$$p(y|z;\psi)=\prod_{t=1}^{n} p(y_t|z_t;\psi). \tag{14}$$

It implies that, given the strengths of attack and defence in $z_1,\ldots,z_n$ and given home ground advantage δ and the dependence coefficient γ, the scores from week to week are conditionally independent. Finally, we can express the joint density as $p(y,z;\psi)=p(y|z;\psi)\,p(z;\psi)$ where

$$p(z;\psi)=p(z_1;\psi)\prod_{t=2}^{n} p(z_t|z_1,\ldots,z_{t-1};\psi). \tag{15}$$

Given the linear Gaussian auto-regressive process for the state vector z_t in expression (11), the evaluation of $p(z_t|z_1,\ldots,z_{t-1};\psi)$ is straightforward. The parameter vector ψ includes the coefficients $\phi_{\kappa,i}$ and $\sigma^2_{\kappa,i}$ for $\kappa=\alpha,\beta$ and $i=1,\ldots,J$. The evaluation of the initial density $p(z_1;\psi)$ can be based on the unconditional properties of z_t. The constants $\mu_{\kappa,i}$, for $\kappa=\alpha,\beta$ and $i=1,\ldots,J$, are incorporated in the initial condition for z_1.

The likelihood function for y is based on the observation density (1) and is given by

$$l(\psi)=p(y;\psi)=\int p(y,z;\psi)\,\mathrm{d}z=\int p(y|z;\psi)\,p(z;\psi)\,\mathrm{d}z, \tag{16}$$

which we want to evaluate for different values of the parameter vector ψ. An analytical solution to evaluate this integral is not available and therefore we rely on numerical evaluation methods. It is well established that numerical integration of a multi-dimensional integral becomes quickly infeasible when the dimension increases. We therefore adopt Monte Carlo simulation methods. We can use such methods since explicit expressions for the densities $p(y|z;\psi)$ and $p(z;\psi)$ are available. A naive Monte Carlo estimate of the likelihood function is given by

$$\hat{l}(\psi)=\frac{1}{M}\sum_{k=1}^{M} p(y|z^{(k)};\psi), \qquad z^{(k)}\sim p(z;\psi), \tag{17}$$

where M is the number of Monte Carlo replications. Since the state vector density $p(z;\psi)$ is associated with the auto-regressive process (11), we obtain $z^{(k)}$ simply via the simulation of auto-regressive processes for a given parameter vector ψ. The draws $z^{(1)},\ldots,z^{(M)}$ are generated independently from each other. This Monte Carlo estimate is numerically not efficient (nor feasible) since the simulated paths have no support from the observed data y. A more effective approach for the evaluation of the likelihood function is to adopt Monte Carlo simulation methods based on importance sampling as proposed by Shephard and Pitt (1997) and Durbin and Koopman (1997). The details of this estimation methodology for likelihood evaluation and for the signal extraction of strengths of attack and defence are discussed in the on-line appendix.

Parameter estimation is carried out via the maximization of the likelihood function with respect to ψ by using standard numerical maximization procedures. To obtain a smooth multi-dimensional likelihood surface in ψ for its maximization, each likelihood evaluation is based on the same random numbers for the generation of M simulated paths of z. The method of maximum likelihood produces parameter estimates with optimal properties in large samples. These optimal properties remain when using Monte Carlo simulation methods whereas the estimates are subject to simulation error.

3. Empirical application

3.1. Data description

We analyse a panel time series of 9 years of football match results from the English Premier League for which 20 football clubs are active in each season. The 20 football clubs that participate in a season vary because the three lowest placed teams at the end of the season are relegated. In the new season they are replaced by three other teams. The number of different teams in the panel is 36. Only 11 teams have played in all nine seasons of our sample and 10 teams have played in only one season. In the time dimension, we span a period from the season 2003–2004 to the season 2011–2012. The seasons run from August to May. Each team plays 38 matches in a season (19 home and 19 away games) so in total we have 380 matches in the season. Most games are played in the afternoons of Saturdays and Sundays; the other games are played during weekday evenings (mostly Mondays). The total number of matches played in our data set is $9 \times 380 = 3420$. The first 7 years are used for parameter estimation and the last 2 years are used to explore the out-of-sample performance of the model. Our data set of football match results can be treated as a time series panel of low counts. In approximately 85% of all matches in our sample, a team has scored only 0, 1 or 2 goals. The distribution of home and away goals scored during the nine seasons is presented in the on-line appendix. Although working with low counts, a significant difference can be identified in the number of goals scored and conceded between the competing teams. A low ranking team rarely scores more than 2 goals in an away match whereas the top ranking teams sometimes reach scores of 5 or higher.

3.2. Details of the basic model

Our analysis of the Premier League football match results is based on the modelling framework that is presented in Section 2. The panel data set has $J = 36$ teams and we therefore need to estimate 36 attack strengths and 36 defence strengths over time; the dimension of the state vector z_t is 72. In comparison with other empirical studies where also state space time series analyses are carried out, the state vector is high dimensional. Since only 20 teams are active during a season, we need to treat large sections of the observations in the time series panel as missing. The state space methodology can treat missing observations in a routine manner; see the discussion in the on-line appendix. The time index t in our analysis does not refer to calendar weeks. Only weeks in a football season for which at least one match is played officially for the Premier League are indexed. The last week of football matches in one season and the first week in the next football season then have consecutive time indices. In our basic model of Section 2, the summer and winter breaks are not treated in our calendar. In Section 2.3 we discuss a modification of our model that accounts for summer and winter breaks. If all teams play their matches weekly, each season consists of 38 weeks. However, owing to unforeseen circumstances, specific matches are postponed and extra time periods need to be added in the data set. The resulting calendar is adopted for the time index t in our analysis.

The dynamic processes of the strengths of attack and defence are given by expression (4) or collectively for the state vector by expression (11). Given the high number of teams, we restrict the auto-regressive coefficients and the disturbance variances to be the same among the teams:

$$\phi_{\alpha,i} = \phi_\alpha, \qquad \phi_{\beta,i} = \phi_\beta, \qquad \sigma^2_{\alpha,i} = \sigma^2_\alpha, \qquad \sigma^2_{\beta,i} = \sigma^2_\beta, \qquad \text{for } i = 1, \ldots, J.$$

These restrictions are not strong since we expect the persistence and the variation of the time varying strengths of attack and defence to be small and similar between the teams. In other words, we expect the strengths of attack and defence for all teams to be evolving slowly over time. However, the strengths of attack and defence of the different teams can still evolve over time by following very different time paths. For the basic model, the home ground advantage δ and the dependence γ are assumed to be the same for all teams and matches. The parameter vector is then given by

$$\psi = (\phi_\alpha, \phi_\beta, \sigma_\alpha^2, \sigma_\beta^2, \delta, \gamma)' \tag{18}$$

and is estimated by the method of Monte Carlo maximum likelihood of Section 2.5. The parameters in ψ are transformed during the estimation process so that the parameter values are within their restrictive ranges, which are

$$0 < \phi_\kappa < 1, \qquad \sigma_\kappa^2 > 0, \qquad \delta > 0, \qquad 0 < \gamma < c,$$

for $\kappa = \alpha, \beta$ and where c represents the upper bound that is implied by the model and derived in the on-line appendix. The transformations for the elements in expression (18) are given by

$$\psi_j = \begin{cases} \dfrac{\exp(\psi_j^*)}{1 + \exp(\psi_j^*)}, & j = 1, 2, \\ \exp(\psi_j^*), & j = 3, 4, 5, \\ \psi_j^*, & j = 6, \end{cases} \tag{19}$$

where ψ_j is the jth element of ψ and ψ_j^* is the transformed coefficient that is actually estimated, for $j = 1, \ldots, 6$. We note that $\psi_6 = \gamma$ is not restricted because the upper bound c is implied by the model.

The signal extraction of the time varying strengths of attack and defence has been carried out by the Monte Carlo methods that were described in Section 2.5. We have used a common set of random numbers to generate M simulated paths for z. The choice of M can be relatively low because we use effective importance sampling methods; the details are provided in the on-line appendix. The computations have been implemented by using the numerical routines that were developed and presented in Koopman *et al.* (2008); they are carried out on a standard computer. We have not encountered numerical problems and the computing times have been relatively low despite the high dimensional state vector.

3.3. Parameter estimates

For our time series panel of number of goals scored by teams in the English Premier League during the seven seasons from 2003–2004 to 2009–2010, the parameter estimates are presented in Table 1. To show the robustness of our Monte Carlo maximum likelihood methods, we present the estimates for various importance sampling replications M. The parameter estimates are robust to different choices of M. We may conclude that the choice of $M = 200$ is sufficient in our analysis but that we can also take $M = 50$ for repeated analyses of the model. Further evidence of the reliability of our results is presented in the on-line appendix.

The estimates of the auto-regressive coefficients of the latent dynamic processes for the signals related to the strengths of attack and defence are close to 1. They imply that the strengths of attack and defence are highly persistent and behave as random-walk processes. However, the auto-regressive coefficients reflect the persistence from week to week during the football seasons for which we do not expect many changes. More changes are expected from season to season in which a season consists of 38 weeks. When we consider the persistence of the signals from season to season, the implied estimates of the auto-regressive coefficients are equal to $0.9985^{38} = 0.94$ and $0.9992^{38} = 0.97$, which still imply persistent processes for the signals but they are stationary.

The estimated disturbance variances for the signals are relatively small, which illustrate that the attack and defence signals do vary over time in a smooth way. We emphasize that the

Table 1. Estimates of parameter vector ψ†

ψ	*Results for $M = 50$*	*Results for $M = 200$*	*Results for $M = 1000$*
ϕ_α	0.9985	0.9985	0.9985
	(0.00044)	(0.00044)	(0.00044)
ϕ_β	0.9992	0.9992	0.9992
	(0.00027)	(0.00027)	(0.00027)
σ_α^2	0.000205	0.000206	0.000206
	(2.20×10^{-5})	(2.27×10^{-5})	(2.28×10^{-5})
σ_β^2	0.000141	0.000143	0.000143
	(2.05×10^{-5})	(2.02×10^{-5})	(2.02×10^{-5})
δ	0.3662	0.3643	0.3641
	(0.0196)	(0.0269)	(0.0252)
γ	0.0966	0.0966	0.0966
	(0.0232)	(0.0232)	(0.0232)
$\hat{l}(\psi)$	−9608.56	−9608.38	−9608.38

†The table reports the Monte Carlo estimates for the parameter vector ψ together with the value of the maximized log-likelihood value for different numbers of simulated paths $M = 50, 200, 1000$. The Monte Carlo estimates of the standard errors are given below the estimates and between parentheses. The data set that was used for estimation covers seven seasons of the English Premier League (from 2003–2004 to 2009–2010).

variances estimated determine the scale of the fluctuations from week to week which we expect to be very small. We do not expect that a top team turns into a relegation candidate during one season. Furthermore, the number of goals in a match scored by one team is typically low. The main changes in the signals for strengths of attack and defence take place in the data over longer time periods.

3.4. *Signal estimates of attack and defence strengths*

By replacing the parameter vector ψ with its estimate as given in Table 1, we can apply the Monte Carlo simulation methods of Section 2.5 to obtain the estimates for the attack and defence signals. The state vector z contains the strengths of attack and defence for all time periods and for all football teams. Once we have computed $\hat{z}$, the importance sampling estimate of the state vector, we can graphically present the estimated attack and defence signals over time together with their standard errors. We note that the standard errors are also computed by using the importance sampling method; see the on-line appendix for details.

The estimation results of the previous section have indicated that the strengths of attack and defence do not fluctuate strongly from week to week but from season to season they can be more substantial. We present in Fig. 1 the signal estimates for the time varying attack and defence strength of the well-known football teams Manchester United and Manchester City. The strength of attack of United has remained relatively constant from 2006 onwards whereas in the earlier years we observe an upward trend in their strength of attack. The strength of attack of City has increased much more dramatically since 2007 and stabilized somewhat in the most recent season of 2011–2012. Manchester City had been able to invest more in high quality players in the previous 5 years owing to the new owners of the club. It is interesting to observe that the investments by City have been more directed towards forward players since

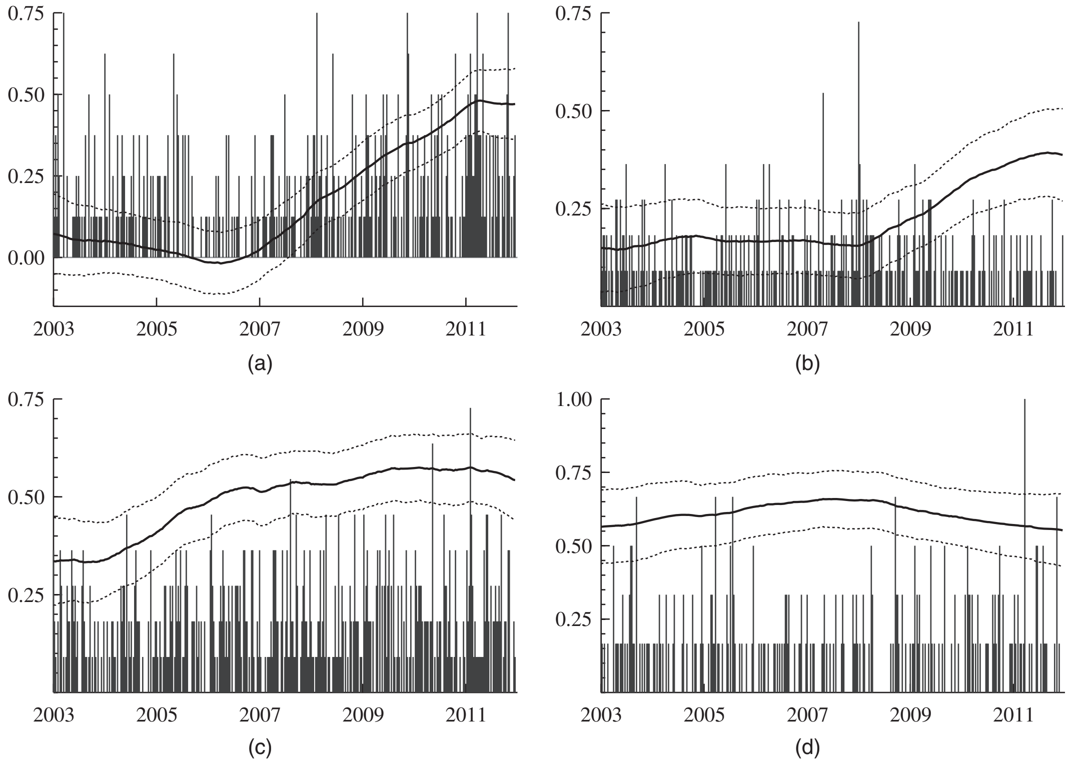

Fig. 1. Strengths of attack and defence of the two highest ranking teams (a), (b) Manchester City and (c), (d) Manchester United at the end of the 2011–2012 season of the English Premier League (———, estimated strength of attack or defence; $\cdots\cdots$, symmetric confidence interval based on 1 standard error; |, number of goals scored and conceded from the 2003–2004 season towards the 2011–2012 season (404 time periods)): (a), (c) strength of attack; (b), (d) strength of defence

the upward trend of the strength of attack is stronger than the trend of the strength of defence. An assessment of the strengths of attack and defence for all teams is presented in the on-line appendix.

3.5. Model evaluations: in sample and out of sample

To validate in-sample estimation and out-of-sample prediction results for the basic model, we present a selection of estimation and testing results for a set of extended, restricted and related model specifications. This study considers the following seven model specifications:

(a) the basic model with parameter estimates presented in Table 1;
(b) the basic model with time invariant strengths of attack and defence (the auto-regressive processes (4) for α_{it} and β_{it} are replaced by fixed coefficients; the state vector (9) reduces to $z_t = \mu$ in expression (11); hence we can adopt the same state space time series analysis but with system matrices $\Phi = 0$ and $H = 0$ in expression (11); the parameter vector consists only of the dependence parameter γ and home ground advantage δ);
(c) the basic model with dependence parameter set equal to 0, i.e. $\gamma = 0$ in expression (1) (the observation model reduces to a double-Poisson distribution for match outcomes);
(d) the basic model with a time varying, team-specific dependence parameter given by

$$\gamma_{ijt} = \gamma^* \sqrt{(\lambda_{x,ijt}\lambda_{y,ijt})}, \qquad \gamma^* \geqslant 0, \tag{20}$$

where γ^* is a scaling coefficient that replaces γ in the parameter vector given in expression (18) (the dependence coefficient is time varying owing to its dependence on the strength of attack and defence; this specification was proposed by Goddard (2005) but the time varying feature of the dependence in expression (20) has not been considered before);

(e) the diagonal inflation model for which the density function (1) is multiplied by expression (6) (the coefficient ω in expression (6) is added to the parameter vector (18));

(f) the basic model with time varying strengths of attack and defence that account for the summer and winter breaks (the disturbance variances for the state vector are time varying as specified in expression (7) and the additional variance parameters are added to the parameter vector (18); here we concentrate on only the long summer break);

(g) the basic model with two home ground advantage parameters, δ_1 for the group {Arsenal, Chelsea, Liverpool, Manchester City, Manchester United} and δ_2 for the group with all other teams; see the discussion in Section 2.3.

3.5.1. In-sample evaluation

For all the model specifications (a)–(g) reported above, we have estimated the parameter vector by the method of Monte Carlo maximum likelihood using the match results in seven seasons of the English Premier League, from 2003–2004 to 2009–2010. Importance sampling methods are used for likelihood evaluation by using a simulation sample size of $M = 50$. The same random draws are used for each model specification, for each parameter vector and for each likelihood evaluation. The usual t-test (for a single restriction) and likelihood ratio statistics are used for the in-sample validation of the restricted and extended model specifications (a)–(g). The test statistics are computed on the basis of maximum likelihood estimates of the parameter vector ψ. Under standard regularity conditions and for sufficiently large sample sizes, the reported t-test and likelihood ratio statistics converge in distribution to a standard normal and a χ^2-distribution with k degrees of freedom, where k is the number of restrictions, respectively. The test statistics are reported in Table 2.

A major aspect of our basic model (a) is the inclusion of time varying strengths of attack and defence. Model (b) reduces the strengths of attack and defence to fixed coefficients. By comparing models (a) and (b) using the likelihood ratio statistic, we conclude that model (b) is not supported by our data set. Another key aspect of our basic model is the use of the bivariate Poisson distribution rather than the double-Poisson distribution as adopted in model (c). The test statistic for model (c) provides clear evidence that our data set favours the model with dependence between the match results. With respect to model (d), we find that the estimated dependence coefficient γ^* in expression (20) is significant. However, the dependence as specified by Goddard (2005) is not strongly favoured in our data set since the maximized likelihood value for basic model (a) is somewhat higher than the maximized likelihood value for model (d).

To account for the overrepresentation of draws in the data set, we consider the diagonal inflation model (e). The maximum likelihood estimate of ω in expression (6) is not significant although the t-test statistic is positive and close to the critical value of 1.96. Hence the number of draws 0–0 and 1–1 that is implied by our basic model is somewhat too small for our data set.

Model (f) allows for breaks in the strengths of attack and defence after the winter and summer holidays in the football calendar. It requires the estimation of four additional variances in the parameter vector. The estimated variances for the winter breaks are not significant and are close to 0. Hence we have re-estimated the model with two additional variances for the summer break only. The two estimated variances have almost equal values. In our final specification

Table 2. Model comparisons: in-sample and out-of-sample results†

Model	*Model specification*	*Number of parameters*	H_0	*Likelihood ratio test*	*t-test*	*Squared loss*	*Diebold–Mariano test*
(a)	Basic model	6				2087.10	
(b)	Basic model with time invariant signals	6/2		123.04‡		2190.80	−3.67‡
(c)	Basic model with no dependence	6/5	$\gamma=0$		4.16‡	2087.90	−0.62
(d)	Basic model with time varying dependence	6/5	$\gamma^*=0$		3.84‡	2088.60	−1.51
(e)	Diagonal inflation model	7/6	$\omega=0$		1.85	2086.70	0.91
(f)	Summer break for signals	7/6	$\sigma^2_{\kappa,\mathrm{S}}=0$		2.84‡	2098.50	−1.75
(g)	Two home ground advantages	7/6	$\delta_1=\delta_2$		0.35	2089.00	−1.08

†We compare the in-sample fit and out-of-sample forecasting accuracy for seven model specifications. The number of parameters (p_1/p_0) is given for each model (p_1) and for the model under the null hypothesis H_0 (p_0); see Section 3.5 for further details. The in-sample results are based on seven seasons of the English Premier League (from 2003–2004 to 2009–2010). The t-tests are computed and presented for the hypotheses with a single restriction whereas the likelihood ratio test is presented for the multiple restriction in model (b). The out-of-sample results are based on the two seasons 2010–2011 and 2011–2012. The squared loss functions and the Diebold–Mariano tests are based on one-step-ahead forecasts from a rolling window sample.
‡Significance at the 5% level of significance.

we therefore restrict the two summer break variances to be equal to each other. The restricted variance estimate is highly significant as indicated by the reported t-test statistic in Table 2. It also affects the estimates of the other variances in the model. In particular, the dynamic coefficients for attack, ϕ_α and σ^2_α, are estimated to be close to 1 and 0 respectively. It implies that the strength of attack is close to a constant within each season and its evolution over time behaves as a step function with breaks at the beginning of each football season. The dynamic coefficients for the strength of defence are not affected in the same way. The strength of defence continues to vary within the season at a slow pace. We present the estimated patterns of attack and defence of Manchester United and Manchester City from model (f) in Fig. 2. We can compare these patterns with those presented in Fig. 1 for our basic model (a). The patterns for a selection of other teams in the English Premier League are presented in the on-line appendix.

Finally, we verify whether the home ground advantage is different for the larger teams in the English Premier League. The home ground advantage parameters δ_1 and δ_2 in model (g) are estimated as an extension of our basic model. The null hypothesis $H_0:\delta_1=\delta_2$ cannot be accepted given the low value of the reported t-test in Table 2. Hence home ground advantage is not significantly different for the larger teams in our data set.

3.5.2. Out-of-sample evaluation

For the out-of-sample evaluation of the models considered, we carry out a one-step-ahead forecasting study. For each model, we forecast the outcome of the matches in the football seasons 2010–2011 and 2011–2012 by using a so-called rolling window strategy. We estimate the parameter vector for the time series of all match results from the seven seasons. At time T, the week before the first week of football season 2010–2011, we forecast the match outcomes for the first week of the season 2010–2011, i.e. time $T+1$, based on a specific model and the

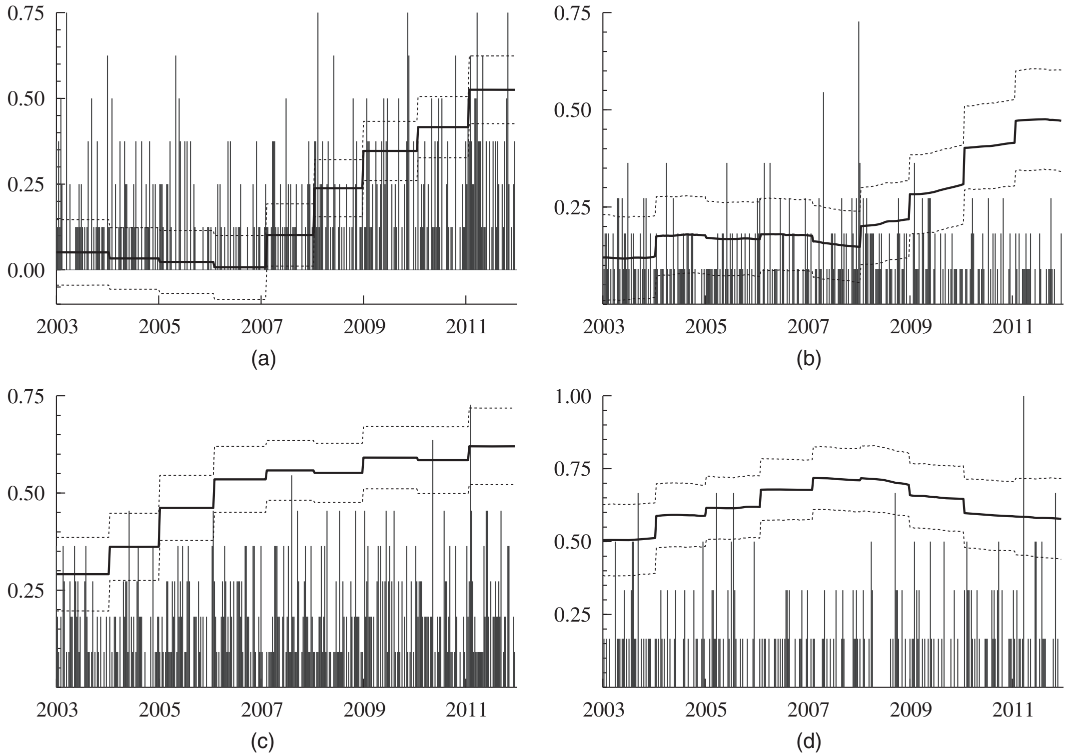

Fig. 2. Strengths of attack and defence of the two highest ranking teams (a), (b) Manchester City and (c), (d) Manchester United at the end of the 2011–2012 season of the English Premier League (the stepwise evolution of the patterns is due to an additional variance for the strengths of attack and defence at the start of the new football season after the summer break) (———, estimated strength of attack or defence; · · · · · ·, symmetric confidence intervals based on 1 standard error; |, number of goals scored and conceded from the 2003–2004 season towards the 2011–2012 season (404 time periods)): (a), (c) strength of attack; (b), (d) strength of defence

estimate of the parameter vector. We then can compare the forecasts with the actual outcomes. The differences between realizations and forecasts are collected in the 20×1 forecast error vector e_{T+1}. Next we compute the sum of squared errors, which we take as our loss function, i.e. $L_{T+1} = e'_{T+1} e_{T+1}$. This loss function is computed for each model, i.e. $L_{t+1}^{(m)}$ for $m = \text{a}, \ldots, \text{g}$. The difference in accuracy compared with our main model can be measured as $d_{T+1}^{(m)} = L_{T+1}^{(\text{a})} - L_{T+1}^{(m)}$ for $m = \text{b}, \ldots, \text{g}$. For the next period $T+1$, we re-estimate the parameter vector by including the match results of time $T+1$ in our data but removing the match results in the first week of our sample, 7 years previously. Hence the estimation sample length remains constant when re-estimating the parameter vector for producing the next forecasts. This procedure of re-estimation and forecasting is then repeated for each week in the two football seasons that we use for our out-of-sample evaluations.

The difference in the one-step-ahead predictions of the models, $d_j^{(m)}$, for $j = T+1, \ldots, T+N$ with out-of-sample length N, are compared with each other on the basis of the Diebold–Mariano (DM) test statistic; see Diebold and Mariano (1995). The test is designed for the null hypothesis of equal out-of-sample predictive accuracy between two competing models. The DM test statistic for model m is computed by

(a) taking the average of the out-of-sample computed values $d_j^{(m)}$ over time, for each $m = \text{b}, \ldots, \text{g}$, and
(b) standardizing this average by a consistent measure of the long-term variance of d_j.

We require the long-term variance because the time series of d_{t+1} is serially correlated by construction since at least only one of the two competing models can be correctly specified. In general, the DM test statistic should not be applied when we compare the predictive accuracy between two nested models since the numerator and denominator of the DM test statistic have their limits at 0, when the in-sample and out-of-sample dimensions increase. However, it is argued by Giacomini and White (2006) that the DM test statistic can still be applied as long as the forecasts are generated with a rolling window and for a relatively short out-of-sample horizon. Diebold and Mariano (1995) showed that the DM test statistic is asymptotically distributed as a standard normal random variable. Hence, we reject the null hypothesis of equal predictive accuracy at the 5% level of significance if the absolute value of the DM test statistic is larger than 1.96. The resulting loss function values and DM test statistics in our out-of-sample forecasting study are reported in Table 2.

The out-of-sample squared loss function values reported in Table 2 show that model (e) in Section 3.5.1 has the smallest loss compared with all other models. Except for models (b) and (f), the forecast losses of the other models are only small and similar in size. This finding is confirmed by the reported DM test statistics which indicate that we cannot reject the hypothesis that any of the models (c), (d), (e) and (g) are equally accurate as model (a) in our out-of-sample forecasting exercise. Although the same conclusion can be drawn for model (f), this model is closest to rejection and appears to provide less accurate forecasts. The stepwise evolution of the strengths of attack and defence from season to season may have a negative effect on its forecasting ability. Given the non-significant DM test statistic for model (c) and despite the in-sample significance of the dependence parameter γ, it appears that the presence of γ does not have much effect on the out-of-sample forecast performance of the basic model. This finding may be due to our choice of a relatively short out-of-sample forecasting window. The only significant DM statistic is reported for model (b) which is consistent with our in-sample rejection of the null hypothesis of time invariant signals. Overall we can conclude that the model extensions of Section 2.3 do not lead to significant improvements in their forecast performance except for model (e). However, the extensions may be more beneficial for longer forecast horizons and for other data sets.

4. Out-of-sample performance in a betting strategy

Finally we verify the out-of-sample performance of our basic model (a) in a realtime study into the betting on a win, a loss or a draw of the home team for a weekly selection of matches during the two seasons of 2010–2011 and 2011–2012. The betting on matches in the English Premier League is immensely popular and is a truly world-wide activity. In our betting evaluation study we carry out the same out-of-sample rolling window strategy as used in the previous section. At time T, we estimate the model parameters and forecast the intensities $\lambda_{x,ij,T+1}$ and $\lambda_{y,ij,T+1}$. The resulting full distributional properties of the next 10 games implied by the bivariate Poisson model (1), with its unknown parameters replaced by their estimates, enables us to compute the probabilities of all possible outcomes of a match. Hence we can compute the probabilities of a win, a loss or a draw, for each match. We can now visit the bookmaker's office and bet on matches accordingly.

Different betting strategies can be pursued and we illustrate our basic and conservative strategy by using an example. Consider the first match of the out-of-sample 2010–2011 season where Aston Villa plays against West Ham. The intensity forecasts are $\lambda_{x,ij,t+1} = 1.7272$ and $\lambda_{y,ij,t+1} = 0.8127$, which correspond to win, loss and draw probabilities for the home team of 0.591, 0.174 and 0.235 respectively. The bookmaker offers the following odds for the home team: 1.96 for a win, 4.03 for a loss and 3.30 for a draw. For each outcome, the expected value of a unit bet on an event A is given by

$$\mathrm{EV}(A) = P(A)\{\mathrm{odds}(A) - 1\} - P(\text{not } A) \times 1 = P(A)\,\mathrm{odds}(A) - 1,$$

where event A represents a win, a loss or a draw of the home team, $P(A)$ is the probability of event A and $\mathrm{odds}(A)$ is the bookmaker's odds for event A. In our illustration we obtain 0.159, -0.300 and -0.224 as expected values for an unit bet on a win, a loss and a draw for the home team respectively. A basic strategy could be to bet on all events for which the expected value is positive, $\mathrm{EV}(A) > 0$. In this illustration we bet on a win for the home team. However, we shall consider a less risky betting strategy which is based on the following guidelines. First, we bet only on 'quality' events which are defined as bets with EVs that exceed some benchmark τ, i.e. $\mathrm{EV}(A) > \tau$ for some $\tau > 0$. Second, we also consider long shot events which are defined as small probability events with very high odds. The probability of losing the bet on a long shot is of course high. We consider events with odds higher than 7 as long shots. Our basic strategy consists of betting a unit value on each quality event for some value of τ. We also bet on long shots but reduce the bet to a fixed value of 0.3 units.

The expected and actual profit for all our bets in the 2010–2011 and 2011–2012 seasons can be determined for a range of τ-values. The sample variance of the computed profit at each time t is obtained by the bootstrap method based on 1000 bootstrap samples; we have carried out a standard bootstrap method as described in Davidson and MacKinnon (2004). The odds for betting are offered by many different bookmakers. We consider the average odds taken from 28–40 bookmakers (depending on the match) which are collected on line at `http://www.football-data.co.uk`. During the two seasons, the 40 bookmakers offered 760 betting opportunities, for all matches played. In the example match between Aston Villa and West Ham, the implied probabilities given by the bookmakers' odds were, on average, 1/1.96, 1/4.03 and 1/3.30 for a win, a loss and a draw respectively. The sum of these probabilities is given by 106.1%. Everything above 100% is the profit of the bookmaker (or the bookmaker's edge) which is 7% on average. This means that the expected profit under random betting of a unit value is -0.07. Random betting is referred to as having a unity bet on a win, a loss or a draw randomly chosen for each match. Hence our betting strategy must achieve an overall return that overtakes the bookmaker's edge of 7% but also generates a positive overall return.

In Fig. 3 we present the outcomes of our betting strategy for various values of τ. In Fig. 3(a) the overall return is presented as the full curve and is compared with the negative overall return of 7%, the bookmaker's edge. The 90% bootstrap confidence interval is represented by the dotted curves. A similar graph was presented by Dixon and Coles (1997). For $\tau = 0$, the majority of betting opportunities is marked by the model as quality bets. For $0 < \tau < 0.12$, the average return is expected to be around 0, which is due to possible model misspecification and parameter uncertainty. We start to obtain positive mean returns at $\tau > 0.12$. The number of betting opportunities becomes small, less than 40, for $\tau = 0.45$. Hence the generated mean returns for $\tau > 0.45$ are not reliable as reflected by the bootstrap confidence intervals. We therefore do not display mean returns for $\tau > 0.45$ in Fig. 3.

We observe that, for small values of τ, the forecasts of our model imply a zero return on average and a negative return on average also finds support in the 90% interval. When the benchmark τ for a quality bet increases, the number of actual bets decreases in our strategy

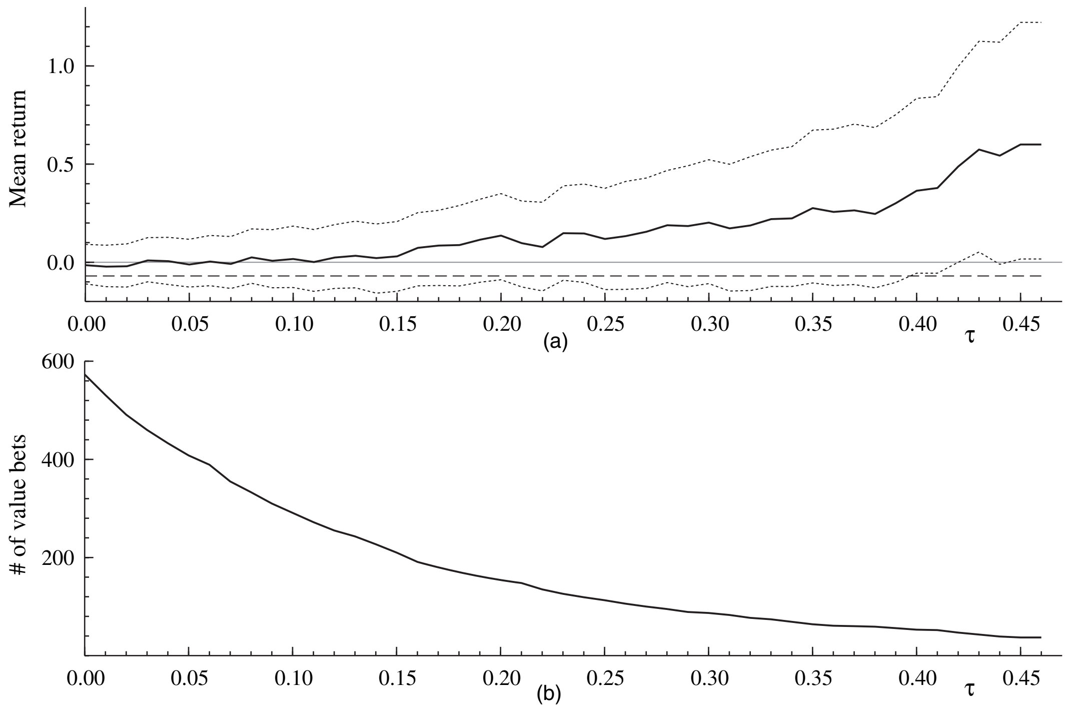

Fig. 3. Returns of a betting strategy for the 2010–2011 and 2011–2012 seasons of the English Premier League: (a) average return from betting on match outcomes by using our strategy for various values of the threshold τ (———), average return under random betting which we have established at -0.07 (– – –) and 90% bootstrap confidence intervals (· · · · · ·); (b) number of quality bets for various values of τ out of the 760 betting opportunities in the two seasons

as is shown in Fig. 3(b). However, the quality bets from a higher benchmark will also provide us with a higher return on average as we learn from Fig. 3(a). When we set $\tau = 0.4$ and play 1 unit for each of the resulting 50 bets in the two seasons, we expect a return of 75 units from the bookmakers: a profit of 25 units or a 50% return, on average.

The average return curve in Fig. 3(a) is not smooth in τ. This is partly due to the role of long shots in this exercise. For example, at $\tau = 0.11$, we have 74 long shots from which eight have been correct, resulting in a net profit of 5.07 units. Even when we bet with 0.3 units for long shots, the betting strategy remains highly variable because, for another value of τ, another small number of correct long shots is obtained that can lead to a very different net profit. A more advanced betting strategy takes into account the variation of odds among the bookmakers. We abstain from such more advanced strategies since we want to illustrate the performance of our model in only a basic and simple betting strategy. The results presented can be used as a benchmark for the more advanced betting strategies based on our model. We regard this validation study as only an example of how our modelling framework can be used in practice.

5. Conclusions

We have presented a non-Gaussian state space model for the analysis and forecasting of football matches. Our basic model takes a match result as a pairwise observation that is assumed to come from a bivariate Poisson distribution with intensity coefficients for the number of goals scored by the two teams and a dependence coefficient for measuring the correlation between the two scores. The intensity coefficients depend on the strengths of attack and defence of the teams and they are allowed to evolve stochastically over time. The intensities are also subject to a fixed coefficient for home ground advantage. The resulting dynamic bivariate Poisson model is a novelty and can be used for the analysis of match results in many different competitions for team sports. Several extensions of the basic model have been considered including amendments for the overrepresentation of draws in data sets, breaks in the strengths of attack and defence after winter and summer breaks, and a team-specific home advantage. Our empirical study is for a data set of match results from nine seasons of the English Premier League. The two seasons of 2010–2011 and 2011–2012 are used as an out-of-sample evaluation period for the forecasting of football match results. The model-based forecasts are of sufficient accuracy for their exploitation in a basic betting strategy. Although we have presented promising results for our basic model and some of its extensions, we believe that further improvements can be made in different directions. First, other dynamic model specifications for the strengths of attack and defence can be considered such as random-walk or long memory processes. Second, our statistical modelling framework uses only match results as data. The forecasting performance of the model can be further improved by adding more information about the matches. For example, potential explanatory variables for match results are the duration between matches played by a team and the travelling distance of the visiting team. Third, our statistical analysis is carried out from a classical perspective. Bayesian Markov chain Monte Carlo methods can be used to obtain predictive densities that account for parameter uncertainty. Fourth, given the popularity of betting on football matches, the odds that are provided by bookmakers are expected to be highly efficient. In such a liquid market of football betting, one can easily find higher odds than the averages that we have used in our study. More advanced betting strategies that take account of the variance of a bet can improve the returns further.

References

Crowder, M., Dixon, M., Ledford, A. and Robinson, M. (2002) Dynamic modelling and prediction of English Football League matches for betting. *Statistician*, **51**, 157–168.

Davidson, R. and MacKinnon, J. G. (2004) *Econometric Theory and Methods*. Oxford: Oxford University Press.

Diebold, F. X. and Mariano, R. S. (1995) Comparing predictive accuracy. *J. Bus. Econ. Statist.*, **13**, 253–265.

Dixon, M. J. and Coles, S. G. (1997) Modelling association football scores and inefficiencies in the football betting market. *Appl. Statist.*, **46**, 265–280.

Durbin, J. and Koopman, S. J. (1997) Monte Carlo maximum likelihood estimation for non-Gaussian state space models. *Biometrika*, **84**, 669–684.

Fahrmeir, L. and Tutz, G. (1994) Dynamic stochastic models for time-dependent ordered paired comparison systems. *J. Am. Statist. Ass.*, **89**, 1438–1449.

Giacomini, R. and White, H. (2006) Tests of conditional predictive ability. *Econometrica*, **74**, 1545–1578.

Glickman, M. E. (2001) Dynamic paired comparison models with stochastic variances. *J. Appl. Statist.*, **28**, 673–689.

Glickman, M. E. and Stern, H. S. (1998) A state-space model for national football league scores. *J. Am. Statist. Ass.*, **93**, 25–35.

Goddard, J. (2005) Regression models for forecasting goals and match results in association football. *Int. J. Forecast.*, **21**, 331–340.

Johnson, N. L., Kotz, S. and Balakrishnan, N. (1997) *Discrete Multivariate Distributions*. New York: Wiley.

Karlis, D. and Ntzoufras, I. (2003) Analysis of sports data by using bivariate Poisson models. *Statistician*, **52**, 381–393.

Karlis, D. and Ntzoufras, I. (2006) Bayesian analysis of the differences of count data. *Statist. Med.*, **25**, 1885–1905.

Karlis, D. and Ntzoufras, I. (2009) Bayesian modelling of football outcomes: using the Skellam's distribution for the goal difference. *IMA J. Mangmnt Math.*, **20**, 133–145.

Knorr-Held, L. (2000) Dynamic rating of sports teams. *Statistician*, **49**, 261–276.

Kocherlakota, S. and Kocherlakota, K. (1992) *Bivariate Discrete Distributions*. New York: Dekker.

Koopman, S. J., Shephard, N. and Doornik, J. A. (2008) *SsfPack 3.0: Statistical Algorithms for Models in State Space Form*. London: Timberlake.
Maher, M. J. (1982) Modelling association football scores. *Statist. Neerland.*, **36**, 109–118.
Ord, K., Fernandes, C. and Harvey, A. C. (1993) Time series models for multivariate series of count data. In *Developments in Time Series Analysis* (ed. T. S. Rao). London: Chapman and Hall.
Owen, A. (2011) Dynamic bayesian forecasting models of football match outcomes with estimation of the evolution variance parameter. *IMA J. Mangmnt Math.*, **22**, 99–113.
Pollard, R. (2008) Home advantage in football: a current review of an unsolved puzzle. *Open Sprts Sci. J.*, **1**, 12–14.
Rue, H. and Salvesen, Ø. (2000) Prediction and retrospective analysis of soccer matches in a league. *Statistician*, **49**, 399–418.
Shephard, N. and Pitt, M. K. (1997) Likelihood analysis of non-Gaussian measurement time series. *Biometrika*, **84**, 653–667.
Skellam, J. G. (1946) The frequency distribution of the difference between two Poisson variates belonging to different populations. *J. R. Statist. Soc.* A, **109**, 296.

Supporting information
Additional ‘supporting information’ may be found in the on-line version of this article:

‘Online Appendix to A dynamic bivariate Poisson model for analysing and forecasting match results in the English Premier League’.