TI 2012-099/III
Tinbergen Institute Discussion Paper


# A Dynamic Bivariate Poisson Model for Analysing and Forecasting Match Results in the English Premier League

*Siem Jan Koopman*[1,2]
*Rutger Lit*[1]

[1] *Faculty of Economics and Business Administration, VU University Amsterdam;*
[2] *Tinbergen Institute.*

Tinbergen Institute is the graduate school and research institute in economics of Erasmus University Rotterdam, the University of Amsterdam and VU University Amsterdam.

More TI discussion papers can be downloaded at http://www.tinbergen.nl

Tinbergen Institute has two locations:

Tinbergen Institute Amsterdam
Gustav Mahlerplein 117
1082 MS Amsterdam
The Netherlands
Tel.: +31(0)20 525 1600

Tinbergen Institute Rotterdam
Burg. Oudlaan 50
3062 PA Rotterdam
The Netherlands
Tel.: +31(0)10 408 8900
Fax: +31(0)10 408 9031

Duisenberg school of finance is a collaboration of the Dutch financial sector and universities, with the ambition to support innovative research and offer top quality academic education in core areas of finance.

DSF research papers can be downloaded at: http://www.dsf.nl/

Duisenberg school of finance
Gustav Mahlerplein 117
1082 MS Amsterdam
The Netherlands
Tel.: +31(0)20 525 8579

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

*Siem Jan Koopman and Rutger Lit*

Department of Econometrics, VU University Amsterdam, NL
Tinbergen Institute, Amsterdam & Rotterdam, NL

September 24, 2012

**Abstract**

Attack and defense strengths of football teams vary over time due to changes in the teams of players or their managers. We develop a statistical model for the analysis and forecasting of football match results which are assumed to come from a bivariate Poisson distribution with intensity coefficients that change stochastically over time. This development presents a novelty in the statistical time series analysis of match results from football or other 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 2010/11 and 2011/12 seasons of the English Premier League. We show that our statistical modeling framework can produce a significant positive return over the bookmaker's odds.

*Some key words*: 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 the next games. Bets can be placed on a win, a loss or a draw but they can also concern the outcome of the match itself. The collection of the predictions are reflected by the bookmaker's odds. In this paper we study a history of all football match results from the English Premier League in the last nine years. The number of goals scored by a team may depend on the attack strength of the team, the defense strength 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 attack and defense strengths of the teams are supposed to 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, 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 double Poisson distribution is used in combination with a dependence parameter which allows for correlation between home and away scores. It represents the phenomenon that the ability or the effort of a team during a match is influenced by the other team or by the way the match progresses. For example, if the home team leads with 1-0, and there is only ten minutes left to play, the away team can become more determined and can take more risk in an effort to end the match in a draw. This possible change in the score due to a change in the behaviour of the team or both teams is captured by a dependence parameter. Furthermore, we let the goal scoring intensities of the two teams depend on the attack and defense strengths of the two teams. The attack and defense strengths for each team is 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 the team to play elsewhere after a number of seasons, it is expected that the attack strength has become weaker of this specific team. In modern football, the composition of a team can be quite different from one season to another season. However, the overall attack and defense strengths are expected to change slowly over time.

The basis of our modelling approach is proposed by Maher (1982). In this study, the double Poisson distribution, with the means expressed as team-specific attack and defense strengths, is adopted as the underlying distribution for goal scoring. Maher (1982) explored the existence of a small correlation between home and away 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, the basic model of Maher is static; the team's attack and defense strengths do not vary over time. Dixon and Coles (1997) consider the double Poisson model with a dependence parameter that is estimated together with the other parameters. They suggest 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 introduce a weighting function to downweight likelihood contributions of observations from the more distant past. Karlis and Ntzoufras (2003) also use a bivariate Poisson distribution; they argue that even a small value for the dependence parameter leads to a more accurate prediction of the number of draws. However, attack and defense strengths are kept static over time in their analysis. Rue and Salvesen (2000) incorporate the framework of Dixon and Coles (1997) within a dynamic generalized linear model and adopt a Bayesian estimation procedure to study the time-varying properties of the football teams. In their empirical analysis, they truncate the number of goals to a maximum of five because they argue that the number of goals beyond five provide no further information about the attack and defense strengths of a team. Crowder, Dixon, Ledford, and Robinson (2002) represent the model of Dixon and Coles (1997) as a non-Gaussian state space

model with time-varying attack and defense strengths. In this study, they use approximate methods for estimation because they state that an exact analysis would be computationally too expensive. Finally, Ord, Fernandes, and Harvey (1993) consider multivariate extensions of a Bayesian dynamic count data model for the analysis and forecasting of number of goals scored by a specific team.

A short overview of the relevant models for our study is presented in Table 1. The bottom right panel of the table is left empty because we believe that our bivariate Poisson model with stochastic time-varying attack and defense strengths is not considered before. We aim to fill this gap in the literature and to show that football match results can be analysed effectively in our statistical modelling framework based on a non-Gaussian state space model where the observed pairs of counts are assumed to come from a bivariate Poisson distribution. We base our analysis on exact maximum likelihood and signal extraction methods which rely on efficient Monte Carlo simulation techniques such as importance sampling. For our statistical analysis, there is no need to truncate the observed match outcomes to some maximum value.

Table 1: Relevant contributions in the literature

An overview of relevant contributions in the literature is given. The two columns provide references to studies in which the attack and defense performances of a team are modelled as a non-stochastic or a stochastic function of time. The two rows provide references to studies in which match results are treated by a bivariate Poisson model with or without a dependence coefficient. The lower right entry is kept empty since we regard this combination as our contribution to the literature.

| | Non-stochastic | Stochastic |
|---|---|---|
| Double Poisson | Maher (1982)<br>Dixon and Coles (1997) | Crowder *et al.* (2002)<br>Rue and Salvesen (2000) |
| Bivariate Poisson | Karlis and Ntzoufras (2003)<br>Goddard (2005) | |

The remainder of the paper is organised 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 attack and defense strengths of a team, and the forecasting of match results. In Section 3 we illustrate the methodology for a high dimensional dataset of football match results from the English Premier League during the seasons from 2003/04 to 2011/12. The first seven seasons are used for parameter estimation and in-sample diagnostic checking of the empirical results while the last two seasons are used for the out-of-sample forecast evaluation of the model. For example, we show that key parts of our model are the dependence coefficient for the correlation between the two scores of a match, home ground advantage and the time-varying attack and defense strengths of the two teams. A forecasting study is presented in Section 4 where we give evidence that our model is capable of turning a positive return over the bookmakers odds by applying a simple betting strategy during the seasons of 2010/11 and 2011/12. Concluding remarks are given in Section 5.

# 2 The 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 is 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

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

for $X = X_{it}$ and $Y = Y_{jt}$, with $\lambda_x$ and $\lambda_y$ being intensity coefficients for $X$ and $Y$, respectively, and $\gamma$ being a coefficient that measures the dependence between the two counts in the pair, $X$ and $Y$. In short notation, we write

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

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

$$\mathbb{E}(X) = \mathbb{V}\text{ar}(X) = \lambda_x + \gamma, \qquad \mathbb{E}(Y) = \mathbb{V}\text{ar}(Y) = \lambda_y + \gamma, \qquad \text{Cov}(X, Y) = \gamma, \quad (2)$$

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

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

The above definition of the bivariate Poisson distribution is not unique, other formulations have also been considered; see, for example, the discussions in Kocherlakota and Kocherlakota (1992) and Johnson, Kotz, and Balakrishnan (1997).

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; see Skellam (1946) and Karlis and Ntzoufras (2003). A particular feature of the Skellam distribution for $X - Y$ is its invariance of $\gamma$ when $(X, Y) \sim BP(\lambda_x, \lambda_y, \gamma)$ for $\gamma > 0$.

## 2.2 Dynamic specification for goal scoring intensities

The scoring intensities of two teams playing against each other are determined by $\lambda_x$, $\lambda_y$ and $\gamma$. In our modelling framework, we let $\lambda_x$ and $\lambda_y$ to 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 played against team $j$, is assumed to depend on the attack strength of team $i$ and the defense strength of team $j$. We also acknowledge the home ground advantage in scoring; this relative advantage is considered to be the same for all teams. The attack strength of the home team $i$ in week $t$ is denoted by $\alpha_{it}$ and its defense strength is denoted by $\beta_{it}$ for $i = 1, \ldots, J$. The home ground advantage is denoted by $\delta$ and is the same for all teams and it is constant over time. The goal scoring intensities for home team $i$ and away team $j$ in week $t$ are then specified as

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

where $\delta$ is the home ground advantage coefficient. We assume that the dependence $\gamma$ between the two scores in a match is the same for all matches played. 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 $\alpha_{it}$'s and defense $\beta_{it}$'s together with coefficient $\delta$, that is $2J+1$ unknowns, when the number of teams is $J > 2$. The time variation of the attack and defense strengths can be identified when we analyse match results from a series of football seasons.

All teams in the competition are assumed to have unique attack and defense strengths which we do not relate to each other. In effect we assume that each team can compose their teams independently of each other. The attack and defense strengths 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 attack and defense strengths as autoregressive processes. We have

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

where $\mu_{\alpha,i}$ and $\mu_{\beta,i}$ are unknown constants, $\phi_{\alpha,i}$ and $\phi_{\beta,i}$ are autoregressive 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

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

for $i = 1, \ldots, J$ and $t = 1, \ldots, n$. The initial conditions for the autoregressive processes $\alpha_{it}$ and $\beta_{it}$ can be based on means and variances of their unconditional distributions which are given by

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

Other, and possibly more complicated, dynamic structures for $\alpha_{it}$ and $\beta_{it}$ can be considered as well but in our current study we will only consider the first-order autoregressive processes as given in (4).

Our basic and simple modelling framework for football match results is introduced. The football match outcomes in our model rely on the stochastic and dynamic properties of the attack ($\alpha_{it}$) and defense ($\beta_{it}$) strengths of the teams but also on the scoring intensity dependence ($\gamma$) of the two teams that play against each other, and the scoring advantage of the home team ($\delta$). The dynamic properties of the attack and defense strengths depend on $\phi_{\kappa,i}$ and $\sigma^2_{\kappa,i}$, for $\kappa = \alpha, \beta$, respectively. Once the paths over time for $\alpha_{it}$ and $\beta_{it}$ are determined, the probability of each possible match outcome can be determined from the bivariate Poisson distribution.

## 2.3 State space representation

For our model-based analysis, it is convenient to present the model into 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

$$(X_{it}, Y_{jt}) \sim BP(\lambda_{x,ijt}, \lambda_{y,ijt}, \gamma), \tag{5}$$

with link functions $\lambda_{x,ijt} = s_{x,ij}(z_t)$ and $\lambda_{y,ijt} = s_{y,ij}(z_t)$ for $i \neq j = 1, \ldots, J$ and with the linear dynamic process for the state vector $z_t$ given by

$$z_t = \mu + \Phi z_{t-1} + \eta_t, \qquad \eta_t \sim \text{NID}(0, H), \tag{6}$$

for $t = 1, \ldots, n$, where $\mu$ is the constant vector, $\Phi$ is the autoregressive coefficient matrix, disturbance vector $\eta_t$ is normally distributed with mean zero and variance matrix $H$, and all unknown coefficients in the model are collected in the parameter vector $\psi$. The state vector $z_t$ contains the attack and defense strengths of all teams, that is

$$z_t = (\alpha_{1t}, \ldots, \alpha_{Jt}, \beta_{1t}, \ldots, \beta_{Jt})', \qquad t = 1, \ldots, n. \tag{7}$$

The initial condition for the state vector $z_1$ can be obtained from the unconditional properties of $z_t$. Similarly, we have $\eta_t = (\eta_{\alpha,1t}, \ldots, \eta_{\alpha,Jt}, \eta_{\beta,1t}, \ldots, \eta_{\beta,Jt})'$. For our modelling framework of the dynamic scoring intensities, it follows that matrices $\mu$, $\Phi$ and $H$ in (6) are given by

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

where $\text{diag}(v)$ refers to a diagonal matrix with the elements of $v$ on the leading diagonal. Given the intensities and the dependence coefficient $\gamma$, we can determine the stochastic

properties of the match result $(X_{it}, Y_{jt})$. The link functions $s_{x,ij}(z_t)$ and $s_{y,ij}(z_t)$ select the appropriate $\alpha_{it}$ and $\beta_{jt}$ elements from $z_t$ and transforms these variables to the intensities as given by (3).

## 2.4 Likelihood function

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 nonlinear 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 number of goals scored by all teams at time $t$ 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

$$p(y_t|z_t;\psi) = \prod_{k=1}^{J/2} BP(\lambda_{x,ijt}, \lambda_{y,ijt}, \gamma), \tag{8}$$

where the index $k$ represents the $k$th match between home team $i$ against visiting team $j$. We notice 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 (3). In this case we can express the signal vector that is implicitly used for the density $p(y_t|z_t;\psi)$ as

$$\mathbb{E}(y_t|z_t;\psi) = \exp(a_t\delta + A_t z_t), \tag{9}$$

where vector $a_t$ has element 1 if the number of goals in the corresponding element of $y_t$ is from the home team and 0 otherwise, matrix $A_t$, with 1s, 0s and -1s, selects the attack strength (+1) of the team, the defense strength (-1) of the opponent team and 0 otherwise. The homeground advantage coefficient $\delta$ is part of the parameter vector $\psi$.

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

$$p(y|z;\psi) = \prod_{t=1}^{n} p(y_t|z_t;\psi). \tag{10}$$

Finally, we can express the joint density as $p(y, z;\psi) = p(y|z;\psi)p(z;\psi)$ where

$$p(z;\psi) = p(z_1;\psi) \prod_{t=2}^{n} p(z_t|z_1, \ldots, z_{t-1};\psi). \tag{11}$$

Given the linear Gaussian autoregressive process for the state vector $z_t$ in (6), the evaluation of $p(z_t|z_1, \ldots, z_{t-1};\psi)$ is straightforward. The parameter vector $\psi$ 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

$$\ell(\psi) = p(y;\psi) = \int p(y,z;\psi)\mathrm{d}z = \int p(y|z;\psi)p(z;\psi)\mathrm{d}z, \tag{12}$$

which we want to evaluate for different values of the parameter vector $\psi$. 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 in practice 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

$$\hat{\ell}(\psi) = \frac{1}{M}\sum_{k=1}^{M} p(y|z^{(k)};\psi), \qquad z^{(k)} \sim p(z;\psi), \tag{13}$$

where $M$ is the number of Monte Carlo replications. Since the state vector density $p(z;\psi)$ is associated with the autoregressive process (6), we obtain $z^{(k)}$ simply via the simulation of autoregressive processes for a given parameter vector $\psi$. 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 are having 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 specific details of this estimation methodology are discussed in the Appendix B.

The maximization of the likelihood function with respect to $\psi$ can then be carried out by standard numerical maximization procedures. To obtain a smooth multi-dimensional likelihood surface in $\psi$ for its maximization, each likelihood evaluation should be based on the same random numbers that generate the series of $M$ simulated paths for $z$. The maximum likelihood method produce parameter estimates with optimal properties in large samples. These optimal properties remain when using Monte Carlo simulation methods appropriately although the estimates are subject to simulation error.

## 2.5 Signal extraction of attack and defense strengths

We use simulation methods for the signal extraction of the attack and defense strengths $\alpha_{it}$ and $\beta_{it}$ in a similar fashion as for the Monte Carlo maximum likelihood estimation of the parameters $\phi_{\kappa,i}$, $\sigma^2_{\kappa,i}$, $\gamma$ and $\delta$, with $i = 1, \ldots, J$, based on the simulated likelihood function (13). However, the same drawbacks apply as for likelihood evaluation via (13). For a given value of the parameter vector $\psi$, we estimate the attack and defense strengths in the state

vector $z$ by evaluating the conditional expectation $\hat{z} = \mathbb{E}(z|y;\psi)$ where

$$\mathbb{E}(z|y;\psi) = \int zp(z|y;\psi)\mathrm{d}z = p(y;\psi)^{-1}\int zp(z,y;\psi)\mathrm{d}z = p(y;\psi)^{-1}\int zp(y|z;\psi)p(z;\psi)\mathrm{d}z.$$

Given the Monte Carlo method for computing the observation density $p(y;\psi)$ and given the known expressions for $p(y|z;\psi)$ and $p(z;\psi)$ above, we can estimate $\hat{z}$ by the same Monte Carlo simulation importance sampling method. This argument can be generalized to the estimation of any known (linear and nonlinear) function of the state vector $z$. It implies that we can evaluate the estimated variance, percentile and distribution of any element of $z$ but also that we can evaluate the estimate of the intensities $\lambda_{x,ijt}$ and $\lambda_{y,ijt}$. Further details are discussed in the Appendix B.

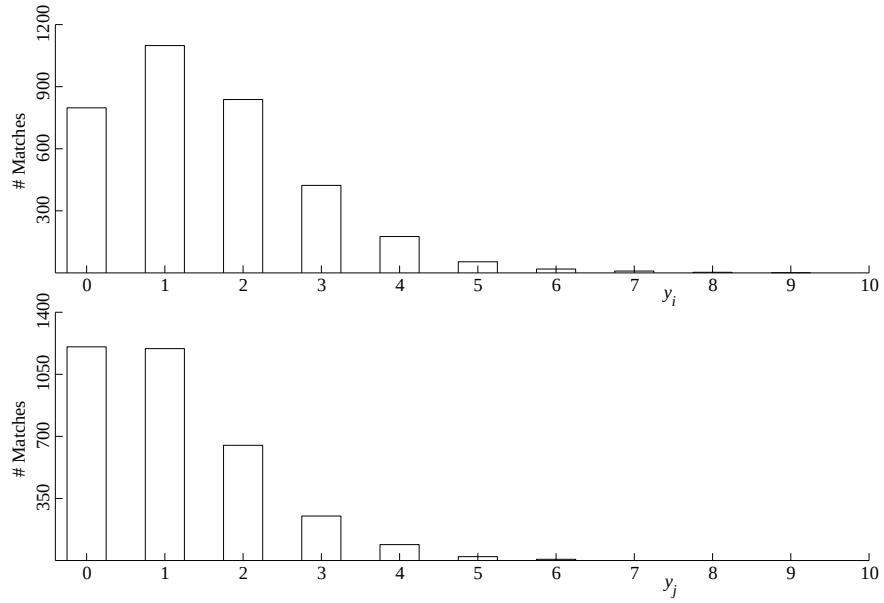
# 3 Empirical application

## 3.1 Data description

We analyse a panel time series of nine 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 while 10 teams have only played in one season. In the time dimension, we span a period from the season 2003/04 to the season 2011/12. The seasons run from August to May. Each team plays 38 matches in a season (19 home and 19 away games) so that 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 dataset is $9 \times 380 = 3,420$. The first seven years are used for parameter estimation and the last two years are used to explore the out-of-sample performance of the model. The data used in our study can be found on `http://www.football-data.co.uk`.

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, the teams have only scored 0, 1 or 2 goals. In Figure 1 we present the distribution of home and away goals scored during the nine seasons. 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 two goals in an away match while the top ranking teams sometimes reach scores of five or higher. This feature of the data is visualized in Figures 2 and 3 where we present the number of goals scored and conceded, respectively, over time for all teams in the data set.

FIGURE 1: HISTOGRAMS OF HOME AND AWAY GOALS



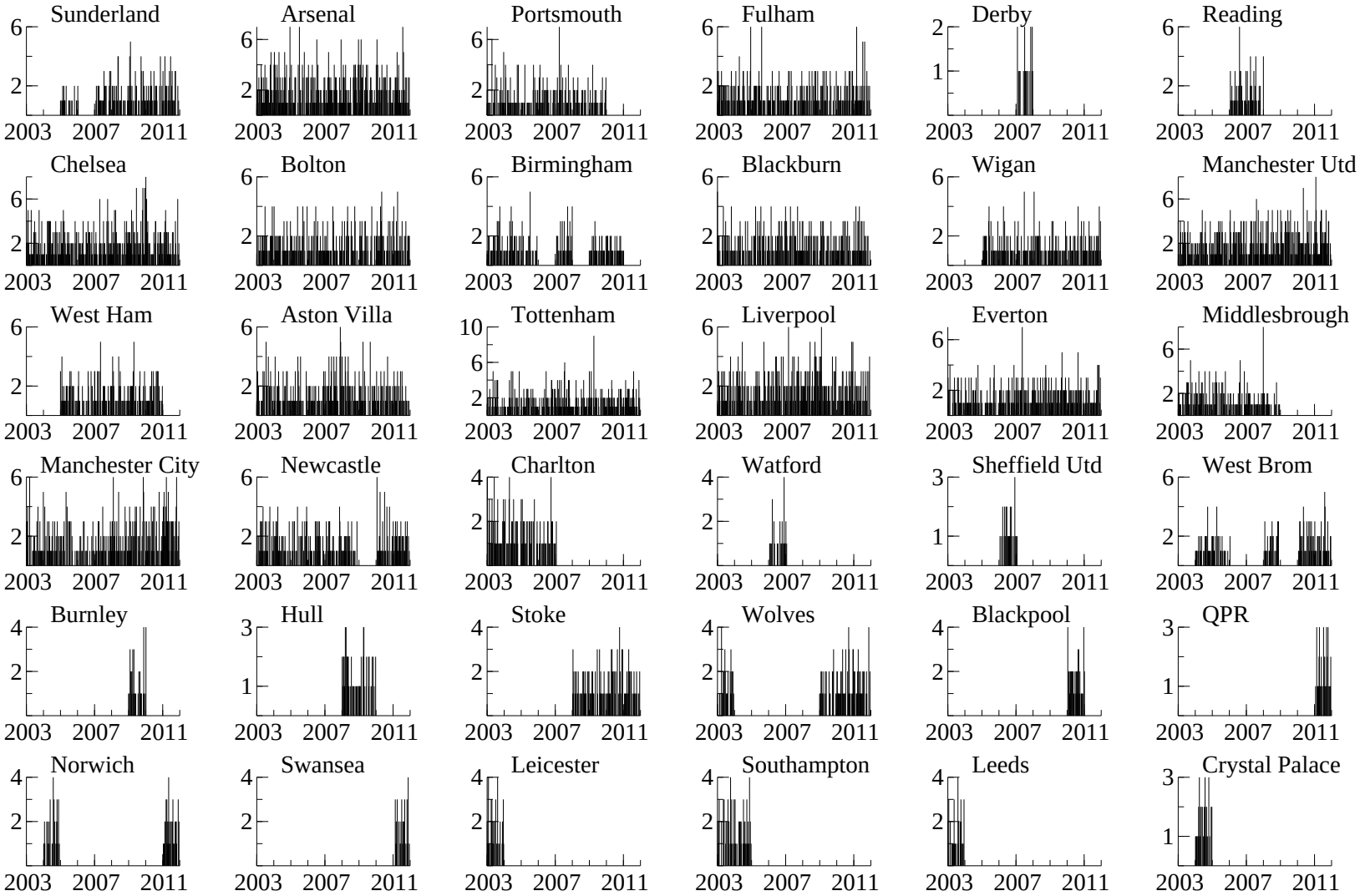

Histograms of home and away goals in the English Premier League over nine seasons ranging from 2003/04 to 2011/12. The average of home goals and away goals is 1.5287 and 1.0994, respectively. Averages are calculated as the average number of goals scored by the home and visiting teams in official time. No matches are played in overtime or finished with penalties.

## 3.2 Model details

Our analysis of the Premier League football match results is based on the modelling framework presented in Section 2. The panel data set has $J = 36$ teams and we therefore need to estimate 36 attack strengths over time and 36 defense strengths; the dimension of the state vector $z_t$ is 72. In comparison to other empirical studies where state space time series analyses are carried out, the state vector has a high dimension. 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 Appendix B. The time index $t$ in our analysis does not refer to calendar weeks. Each week in a football season when at least one match is played officially for the Premier League is indexed and is indicated with some time index $t$. The last week of football matches in one season and the first week in the next football season have then consecutive time index values. It means that summer (but also winter) breaks are not accounted for in our analysis. If all teams play their matches on a weekly basis, each season consists of 38 weeks. However, due to unforeseen circumstances, specific matches are postponed and extra time periods need to be added in the data set. The resulting time index $t$ is adopted in our analysis; see also Figures 2 and 3.

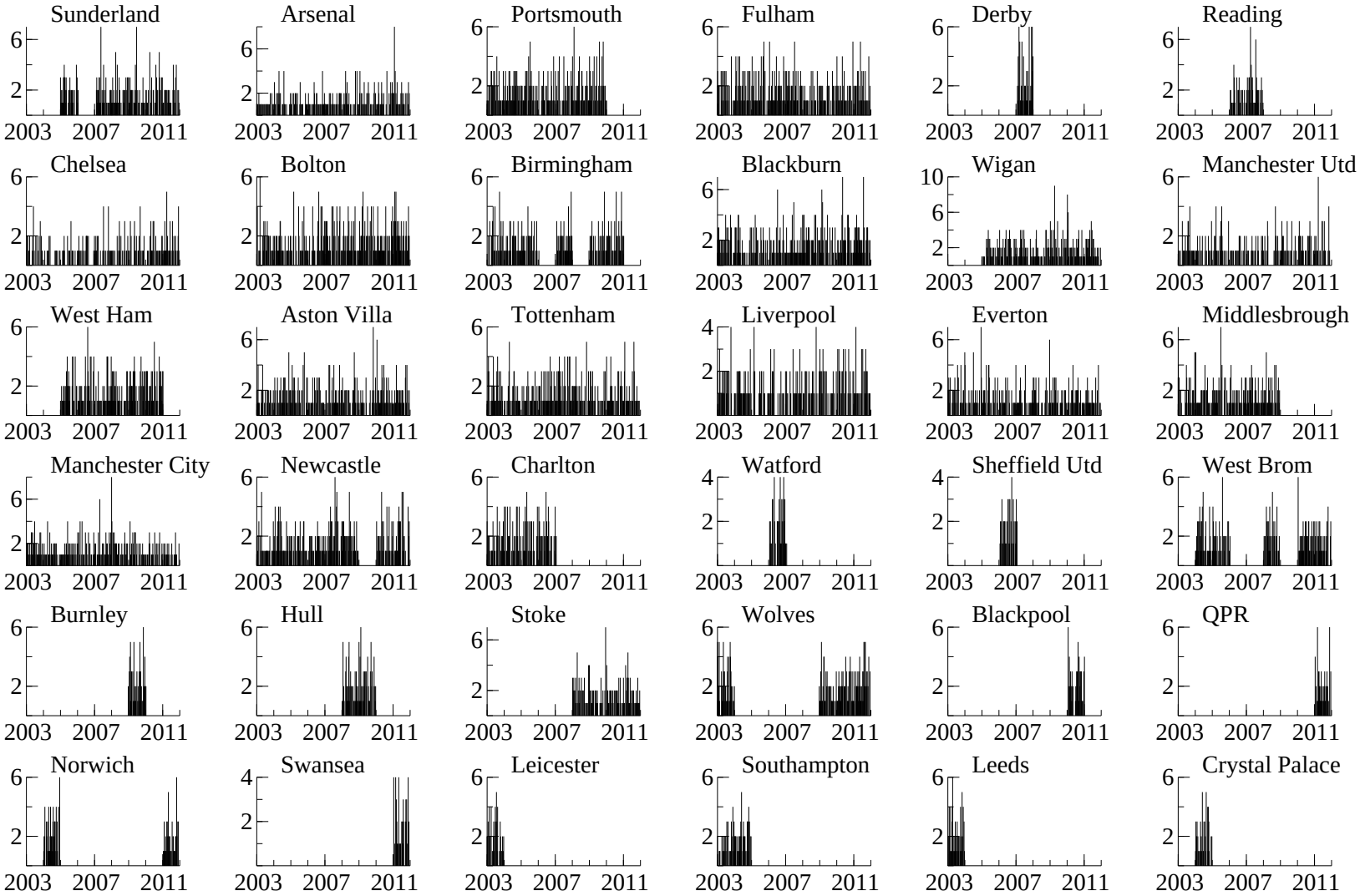
The dynamic properties of the attack and defense strengths are given by (4) or collectively in the state vector by (6). Given the high number of teams, we restrict the autoregressive

Figure 2: Goals scored over time




The number of goals scored by football teams from the 2003/04 towards the 2011/12 season of the English Premier League. Data is given in transaction time meaning that the years on the $x$-axis are football years and not calendar years. Due to promotion and relegation, some teams only compete for one year in the Premier League while other teams are active in all seasons.

Figure 3: Goals conceded over time




The number of goals conceded over time by football teams from the 2003/04 towards the 2011/12 season of the English Premier League. Data is given in transaction time meaning that the years on the $x$-axis are football years and not calendar years. Due to promotion and relegation, some teams only compete for one year in the Premier League while other teams are active in all seasons.

coefficients and the disturbance variances to be the same amongst the teams:

$$\phi_{\alpha,i} = \phi_\alpha, \qquad \phi_{\beta,i} = \phi_\beta, \qquad \sigma^2_{\alpha,i} = \sigma^2_\alpha, \qquad \sigma^2_{\beta,i} = \sigma^2_\beta,$$

for $i = 1, \ldots, J$. These restrictions are not strong since we expect the persistence and the variation of the time-varying attack and defense strengths to be small and similar amongst the teams. In other words, we expect the attack and defense strengths for all teams to be evolving slowly over time. The attack and defense time paths for all teams can still change very differently over time. The home ground advantage $\delta$ in (4) is also restricted to be the same for all teams. We also do not regard this restriction as being strong. When a home team plays on an artificial pitch, it would clearly be a different matter; see the study of Barnett and Hilditch (1993). In such cases we could add an extra parameter in the model to account for this effect. However, artificial pitches are prohibited in the English Premier League and therefore the issue does not arise. Finally, the dependence in the scoring intensities of two opposing teams is $\gamma$ and appears in (1). This dependence parameter is the same for all matches played.

The parameter vector is then given by

$$\psi = (\phi_\alpha, \phi_\beta, \sigma^2_\alpha, \sigma^2_\beta, \delta, \gamma)',$$

and is estimated by the method of Monte Carlo maximum likelihood as described in Section 2.4. The parameters are transformed during the estimation process such that the parameter values are within their restrictive ranges as given by

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

for $\kappa = \alpha, \beta$ and where $c$ represents the upper bound given in Appendix A. The signal extraction of the time-varying attack and defense strengths has been carried out by the Monte Carlo methods as described in Section 2.5. We have used a common set of random numbers to generate $M$ Monte Carlo paths for $z$. The choice of $M$ can be relatively low due to the effective importance sampling methods that have been used; the details are provided in the Appendix B. The computations have been implemented using the numerical routines developed and presented in Koopman, Shephard, and Doornik (2008); they are carried out on a standard computer. We have not encountered numerical problems while the computing times have been relatively short 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 2003/04 to 2009/10, the parameter estimates are presented in Table 2. To show the robustness of our Monte Carlo maximum likelihood methods, we present the estimates for different importance sampling replications $M$.

The parameter estimates are clearly robust to different choices of $M$. We may conclude

Table 2: Estimates of parameter vector $\psi$

The table reports the Monte Carlo estimates for the parameter vector $\psi$ together with the value of the maximized loglikelihood value for number of simulated paths $M = 50, 200, 1000$. The Monte Carlo estimates of the standard errors are between parentheses. The dataset used for estimation consists of seven seasons of the English Premier League (2003/2004 – 2009/2010).

| $\psi$ | $M = 50$ | $M = 200$ | $M = 1,000$ |
|---|---|---|---|
| $\phi_\alpha$ | 0.9985 | 0.9985 | 0.9985 |
| | (0.00044) | (0.00044) | (0.00044) |
| $\phi_\beta$ | 0.9992 | 0.9992 | 0.9992 |
| | (0.00027) | (0.00027) | (0.00027) |
| $\sigma_\alpha^2$ | 0.000205 | 0.000206 | 0.000206 |
| | (2.20e-05) | (2.27e-05) | (2.28e-05) |
| $\sigma_\beta^2$ | 0.000141 | 0.000143 | 0.000143 |
| | (2.05e-05) | (2.02e-05) | (2.02e-05) |
| $\delta$ | 0.3662 | 0.3643 | 0.3641 |
| | (0.0196) | (0.0269) | (0.0252) |
| $\gamma$ | 0.0966 | 0.0966 | 0.0966 |
| | (0.0232) | (0.0232) | (0.0232) |
| $\hat{\ell}(\psi)$ | -9608.56 | -9608.38 | -9608.38 |

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. Since we only need to consider small Monte Carlo simulation samples, the computing times are relatively short. Further evidence of the reliability of our results is discussed in Appendix B.

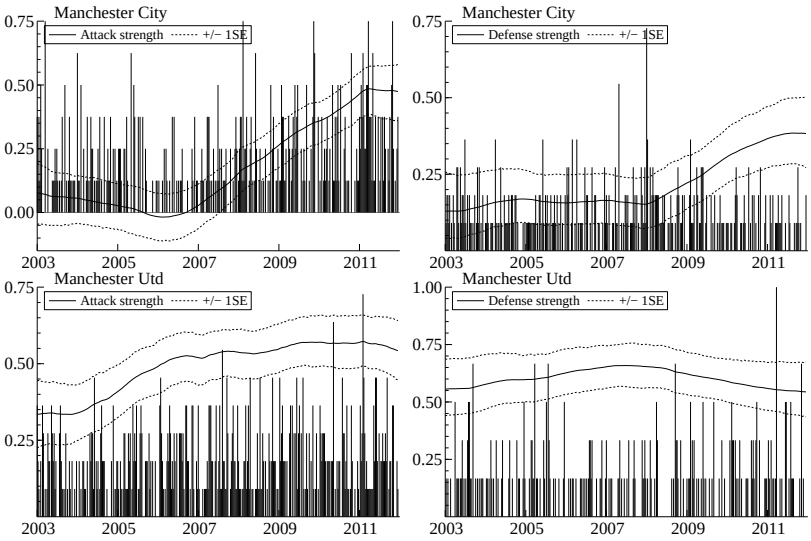
Table 2 presents the estimates of the autoregressive coefficients of the latent dynamic processes for the signals related to the attack and defense strengths. Apparently, the coefficients are estimated close to one which indicate that the attack and defense strengths are highly persistent and behave almost as random walks. However, it reflects the persistence from week to week during the football seasons during which we do not expect much 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 we obtain autoregressive coefficients equal to $(0.9985)^{38} = 0.94$ and $(0.9992)^{38} = 0.97$ which still imply persistent processes for the signals but their behaviour are clearly stationary.

The estimated disturbance variances for the signals are reported as relatively small values which illustrate that the attack and defense signals do not vary much over time. We again emphasize that these estimated variances 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. In our data set, 85% of the scores consists of counts of $0, 1$ or 2. Hence changes in the signals for attack and defense strengths can only be observed from the data in a very subtle way.

## 3.4 Signal estimates of attack and defense strengths

By replacing the parameter vector $\psi$ with its estimate as given in Table 2, we can apply the Monte Carlo signal extraction method of Section 2.5 to obtain the estimates for the attack and defense signals. The state vector $z$ consists of all these signals for all time periods and for all football teams. Once we have computed its importance sampling estimate $\hat{z}$, we can present elements of these estimates over time together with their standard errors. The standard errors can also be computed by the importance sampling method as indicated in Section 2.5.

FIGURE 4: ATTACK AND DEFENSE STRENGTHS OF TWO HIGH RANKING TEAMS

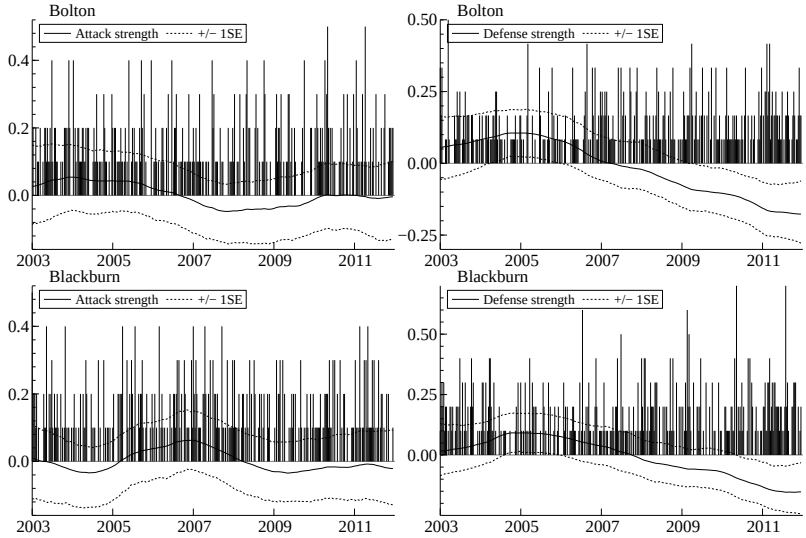



The panels show attack and defense strengths of the two highest ranking teams at the end of the 2011/12 season of the English Premier League. The solid lines are the estimated attack and defense strengths. The dotted lines provide the symmetric confidence intervals based on one standard error. The bars represent the number of goals scored and conceded from the 2003/04 towards the 2011/12 season which accounts for 404 time periods.

The estimation results of the previous section have indicated that the attack and defense strengths do not fluctuate strongly over time, from week to week. However, the changes in attack and defense strengths from season to season can be more substantial. We present in Figures 4 the signal estimates for the time-varying attack and defense strength of the well-known football teams Manchester United and Manchester City. The attack strength of United have remained relatively constant from 2006 onwards while in the earlier years we observe an upwards trend in their attack strength. The attack strength of City has increased much more dramatically since 2007 and stabilised somewhat in the most recent season of 2011/12. Manchester City has been able to invest more in high quality players in the last five years due to the new owners of the club. It is interesting to observe that the investments of Manchester City has been more directed towards forward players since the upward trend of the attack strength is stronger than the upward trend of the defense strength.

The estimated attack and defense strengths for Bolton and Blackburn are presented in

FIGURE 5: ATTACK AND DEFENSE STRENGTHS OF TWO LOW RANKING TEAMS




The panels show attack and defense strengths of the two lowest ranking teams at the end of the 2011/12 season of the English Premier League. The solid lines are the estimated attack and defense strengths. The dotted lines provide the symmetric confidence intervals based on one standard error. The bars represent the number of goals scored and conceded from the 2003/04 towards the 2011/12 season which accounts for 404 time periods.

Figure 5. These two football teams from the Premier League have been low ranking teams at the end of the 2011/12 season. The defense strength of Bolton has deteriorated significantly in the last years and hence it may explain their low ranking. For both teams the attack strengths have remained stable over the years. Hence the model may suggest that both teams should invest more in their defense strengths in the coming years.

## 3.5 Model validation: in-sample and out-of-sample

To validate in-sample estimation and out-of-sample prediction results for our main model, we present a selection of estimation and testing results for a set of restricted and related model specifications. Based on these results and comparisons, we can investigate the empirical relevance or the contribution of the different features of our main model. The comparative study includes six different model specifications and they are listed and labelled below.

($a$) Main model with the estimation results presented in Table 2 and discussed in the previous section.

($b$) Main model with dependence parameter set equal to zero, that is $\gamma = 0$. This model specification reduces the observation density to the one of the independent double Poisson distribution.

($c$) Main model with the dependence parameter equal to the team-dependent and time-varying specification

$$\gamma_{ijt} = \gamma^* \sqrt{\lambda_{x,ijt}\lambda_{y,ijt}}, \qquad \gamma^* \geq 0, \tag{14}$$

where $\gamma^*$ is a scaling coefficient that we estimate together with the other unknown parameters. The dependence coefficient is time-varying due to its dependence on the time-varying attack and defense strengths. This model specification of the dependence parameter is proposed by Goddard (2005) but the time-varying feature has not been considered in his study.

($d$) Main model where the goal scoring intensities are not modelled as (3) but reduced to $\lambda_{it} = \exp(\theta_{it})$ for all $i = 1, \ldots, J$ with the same autoregressive specifications as in (4) but with $\alpha_{it}$ replaced by $\theta_{it}$ and with $\beta_{jt}$ deleted. The decomposition of the log-intensities of goals scored into attack and defense strengths of teams is aborted. The home ground advantage parameter $\delta$ also does not play a role in this specification. The dependence parameter $\gamma$ is still estimated.

($e$) Main model with time-invariant attack and defense strengths. The autoregressive processes (4) are dropped for $\alpha_{it}$ and $\beta_{it}$ and we take them as fixed coefficients in the state vector (7), that is $z_t = \mu$ in (6). We can adopt the same state space time series analysis but with system matrices $\Phi = 0$ and $H = 0$ in (6). The dependence parameter $\gamma$ only needs to be estimated.

($f$) Model specification ($d$) but with time-invariant attack and defense strengths as in ($e$).

For all these model specifications, $(a), \ldots, (f)$, the loglikelihood value is calculated as described in Section 2.4 using match results in seven seasons of the English Premier League, those from 2003/04 to 2009/10. For each model specification, the loglikelihood function is maximised with respect to the unknown parameters. The application of the importance sampling method for the Monte Carlo evaluation of the loglikelihood function is based on a simulation sample size of $N = 50$. The same random draws are used for the evaluation of the loglikelihood value, for each model specification and for each value of the parameter vector. For our in-sample validation of restrictions imposed on our main model specification ($a$), we adopt the likelihood ratio (LR) test statistic as given by $LR = -2\left[\ell_{(b)}(\hat{\psi}_{(b)}) - \ell_{(a)}(\hat{\psi}_{(a)})\right]$ where $\ell_{(m)}(\psi_{(m)})$ is the loglikelihood function for model $(m)$ and $\hat{\psi}_{(m)}$ is the maximum likelihood estimate of the parameter vector $\psi_{(m)}$ for model $(m)$ with $m = a, b$. Under standard regularity conditions and as the sample size increases, the LR test converges in distribution to a $\chi^2$ with $k$ degrees of freedom where $k$ is the number of elements that vector $\psi_{(a)}$ exceeds vector $\psi_{(b)}$. The maximised loglikelihood values and the LR test statistics are reported in Table 3 for all models. In terms of the LR tests, we can conclude that all features of our main model cannot be rejected by the restrictions or simplifications implied by the models ($b$) upto ($f$). The exception may be model ($c$) where the alternative specification of Goddard (2005) is close to the maximised loglikelihood value of model ($a$) but in actual levels, the likelihood value of our main model is higher. Hence we do not feel that sufficient evidence is given for the incorporation of this specification into our main model.

The rejection of the hypothesis $\gamma = 0$ as implied by model ($b$), also implies the rejection of the double Poisson model. It confirms the in-sample significance of our estimate for $\gamma$ as

reported in Table 2. Earlier contributions have reported that the independent double Poisson model has a tendency to underpredict the number of draws in a competition; see, for example, Dixon and Coles (1997) and Karlis and Ntzoufras (2003). In the latter article, the importance of a relatively small value for the dependence parameter $\gamma$, implying a weak correlation between home and away scores, is illustrated by a simulation exercise. For example, when they set $\gamma = 0.05$ with $\lambda_x = \lambda_y = 1$ in (1), the number of draws increased by 3.3% compared to the double Poisson model, that is $\gamma = 0$. When the dependence is set to $\gamma = 0.20$, the number of draws shows an increase of 14%. In our study, $\gamma$ is estimated close to 0.10 and we may therefore conclude that our main model shows an increase of more than 6.5% in the number of expected draws, when compared to model $(b)$.

The strong in-sample rejections of the models $(d)$, $(e)$ and $(f)$ imply that the modelling framework of Maher (1982) and the time-varying nature of the attack and defense strengths are clear features in the football match results data from the English Premier League. It provides further support to the empirical in-sample results which are reported and discussed in the previous subsections.

For the out-of-sample validation of our main model, we carry out a thoroughly conducted one-step ahead forecasting study. For each model, we forecast the outcome of the matches in the football seasons 2010/11 and 2011/12 using a so-called rolling window strategy. We have estimated the parameter vector for the multiple time series of seven seasons of match results. At time $t$, the week before the first week of football season 2010/11, we forecast the match outcomes for the first week of the season 2010/11, that is time $t + 1$, based on our model and the estimated parameter vector. Since the realisations are known, we can compare the forecasts with the actual outcomes. The differences between realisations and forecasts are collected in the $20 \times 1$ forecast error vector $e_{t+1}$. Next we compute the sum of squared errors which we take as our loss function, that is $L_{t+1} = e'_{t+1} e_{t+1}$. This loss function is computed for each model, that is $L^{(m)}_{t+1}$ for $m = a, \ldots, f$. The difference in accuracy compared to our main model can be measured as $d^{(m)}_{t+1} = L^{(a)}_{t+1} - L^{(m)}_{t+1}$ for $m = b, \ldots, f$. 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, seven years ago. Hence the estimation sample 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 validation. The predictive accuracies of the different models 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 (i) taking the average of the out-of-sample computed values $d^{(m)}_{t+1}$'s over time, for each $m = b, \ldots, f$; (ii) standardizing this average by a consistent measure of the long-term variance of $d_{t+1}$. 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 zero, 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) show 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% significance level if the absolute value of the DM test statistic is larger than 1.96. Our main model produces the most accurate out-of-sample forecasts in comparison to a rival model when the DM value is smaller than $-1.96$. The resulting loss function values and DM test statistics in our out-of-sample forecasting study are reported in Table 3.

TABLE 3: MODEL COMPARISONS: IN-SAMPLE AND OUT-OF-SAMPLE RESULTS

We compare the in-sample fit and out-of-sample forecasting accuracy for six competing model specifications. The maximized loglikelihood values and the likelihood-ratio (LR) tests are computed by importance sampling methods with $M = 50$ simulation draws. The in-sample results are based on seven seasons of the English Premier League (from 2003/04 to 2009/10). The squared loss functions and the Diebold-Mariano (DM) tests are based on one-step ahead forecasts from a rolling window sample. The out-of-sample results are based on the two seasons 2010/11 and 2011/12. The test statistic values with $^{**}$ indicate significance at the 5% significance level.

| Model | Restrictions | #pars | log lik | LR test | sqr loss | DM test |
|---|---|---|---|---|---|---|
| $(a)$ | None | 6 | $-9608.56$ | | 2088.40 | |
| $(b)$ | $\gamma = 0$ | 5 | $-9617.44$ | $17.76^{**}$ | 2089.90 | $-0.63$ |
| $(c)$ | $\gamma$ as (14) with $\hat{\gamma}^* = 0.0812$ | 6 | $-9609.68$ | | 2090.20 | $-1.35$ |
| $(d)$ | $\delta = 0$, only intensity signal | 3 | $-9851.43$ | $485.74^{**}$ | 2249.00 | $-4.34^{**}$ |
| $(e)$ | time-invariant signals for $(a)$ | 1 | $-9670.08$ | $123.04^{**}$ | 2189.10 | $-3.49^{**}$ |
| $(f)$ | time-invariant signal for $(d)$ | 1 | $-9884.93$ | $67.01^{**}$ | 2272.80 | $-4.65^{**}$ |

The out-of-sample squared loss function values reported in Table 3 show that model $(a)$ has the smallest loss compared to the other five models. However, the losses for models $(b)$ and $(c)$ are also small and close to the loss of model $(a)$. It appears that the dependence parameter $\gamma$ does not have a large impact on the out-of-sample forecast accuracy of the model while the $\gamma$ has been estimated as strongly significant in-sample. A possible explanation is the relatively short out-of-sample that we have used in our study. The results are confirmed by the reported DM test statistics which indicate that we cannot reject the hypothesis that models $(b)$ and $(c)$ are equally accurate as model $(a)$ in out-of-sample forecasting. The in-sample rejections of models $(d)$, $(e)$ and $(f)$ relative to model $(a)$ are confirmed by the out-of-sample statistics in Table 3. The loss functions for these models are much higher and the DM test statistics show that the equal predictive accuracy hypothesis can be rejected for the models $(d)$, $(e)$ and $(f)$ when compared to model $(a)$. Overall we conclude that model $(a)$ is our preferred specification for both in-sample fit and out-of-sample forecasting accuracy for the range of specifications considered here.

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

We have shown that our statistical dynamic modelling framework is able to forecast match results accurately in comparison to other specifications. It is therefore interesting to verify the out-of-sample performance of our model for the betting on a win, a loss or a draw of the home team for a selection of matches each week during the two seasons of 2010/11 and 2011/12. The betting on matches in the English Premier League is immense popular and is really a 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 $\gamma$ and the other parameters and we forecast the intensities $\lambda_{x,ij,t+1}$ and $\lambda_{y,ij,t+1}$ using the data upto time $t$. We then have the full distributional properties of the next ten games implied by the bivariate Poisson model (1) with its unknown parameters replaced by their estimates and forecasts. It enables us to compute the probabilities of all possible outcomes of a match and hence the probabilities of a win, a loss or a draw for the home team. We compute these probabilities based on the Skellam distribution; see Section 2.1. Although the Skellam distribution is invariant to $\gamma$, the dependence coefficient remains to affect the estimated properties of the attack and defense strengths. Once the probabilities for a win, a loss or a draw (they sum up to one) are established for all ten next week's matches, we can visit the bookmaker's office and bet on these matches.

Different betting strategies can be followed and we illustrate our basic and conservative strategy as follows. For example, consider the first match of the out-of-sample 2010/2011 season where Aston Villa played against West Ham. The forecasted intensities for this match 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 (EV) of a unity bet on an event $A$ is given by

$$\text{EV}(A) = P(A) \times [\text{Odds}(A) - 1] - P(\text{Not } A) \times 1 = P(A) \times \text{Odds}(A) - 1,$$

where event $A$ represents a win, a loss or a draw of the home team, $P(A)$ is the probability of event $A$ and $\text{Odds}(A)$ 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 unity 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, $\text{EV}(A) > 0$. In this illustration we then bet on a win for the home team. However, we will 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 $\tau$, that is $\text{EV}(A) > \tau$ for some $\tau > 0$. Second, we also consider possible longshot events which are defined as small probability events with such high odds that they pass as quality events. The probability of losing the bet on a longshot is of course high. We therefore explicitly mark longshots in our study. We consider events with odds higher than 7 as longshots.

Our validation exercise is for the English Premier League data set. The forecasts of

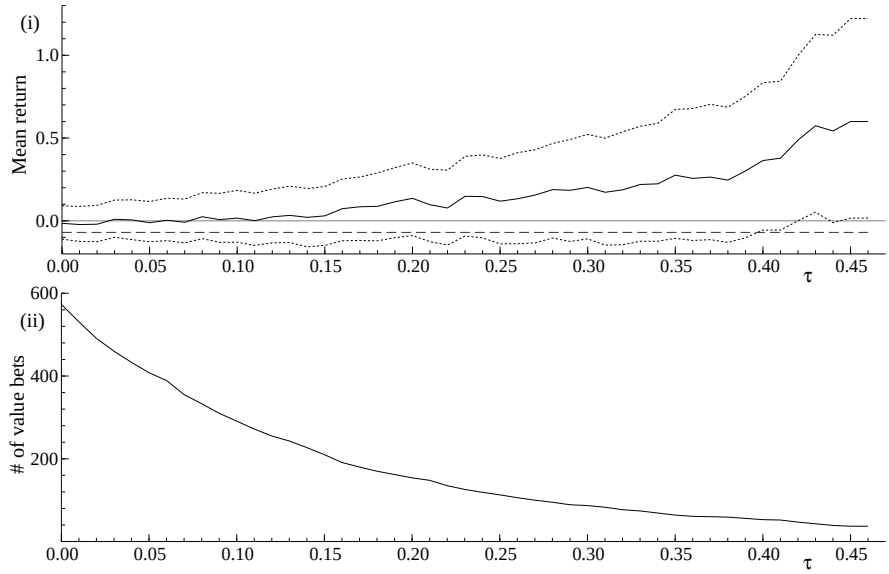
the intensities and hence the forecasts of the probabilities of all possible match results (the events of a win, a loss and a draw) are computed by the rolling window strategy used for our out-of-sample validation in the previous section. We follow the basic strategy as described above and we bet an unity value on each quality event for some value of $\tau$. We also bet on longshots but reduce this bet to a value of 0.3 unit.

The expected and actual profit for all our bets in the 2010/11 and 2011/12 seasons can now be determined as described above for a range of $\tau$ values. The sample variance of the computed profit at each time $t$ is obtained by the bootstrap method based on $1,000$ 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 to 40 bookmakers (depending on the match) which are collected online at `http://www.football-data.co.uk`. During the two seasons, the 40 bookmakers have provided us with 760 betting opportunities, for all matches played. In the example match between Aston Villa and West Ham above, the implied probabilities given by the bookmakers odds have been, on average, 1/1.96, 1/4.03 and 1/3.30 in the respective order of a win, a loss and a draw by the home team. The sum of these probabilties is given by 106.1%. Everything above the 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 unity value is $-0.07$. Random betting is referred to as having an 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 Figure 6 we present the outcomes of our betting strategy for different values of $\tau$. In the first panel $(i)$ the overall return is presented as the solid line and it is compared with the negative overall return of 7%, the bookmaker's edge. The 90% bootstrap confidence interval is represented by the dotted lines. A similar graph is 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 zero 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 become small, less than 40, for $\tau = 0.45$. Hence the generated mean returns for $\tau > 0.45$ are not reliable which is reflected by the bootstrap confidence intervals. We do not report the mean returns for $\tau > 0.45$ in Figure 6.

We observe that for small values of $\tau$, the forecasts of our model imply a zero return on average while a negative return on average also finds support in the 90% interval. When the benchmark $\tau$ for a quality bet increases, the number of actual bets decreases in our strategy as is shown in panel $(ii)$ of Figure 6. However, the quality bets from a higher benchmark will also provide us with a higher return on average as we learn from panel $(i)$. For example, when we set $\tau$ equal to 0.40, we take 50 bets in the two seasons and we expect a return of just below 0.5 on average. When we then play with 1 unit for each of the 50 bets, we expect to receive 75 units from the bookmakers in return; this is a profit of 25 units, a 50% return, on average. Since negative returns are not likely given the 90% confidence interval, we do not expect to loose money in our betting strategy for $\tau = 0.4$.

FIGURE 6: RETURNS OF BETTING STRATEGY FOR THE 2010/11 AND 2011/12 SEASONS




(i) The solid line is the average return from betting on match outcomes in the 2010/11 and 2011/12 seasons of the English Premier League using our strategy for different values of the threshold $\tau$; the dashed line represents the average return under random betting which we have established at $-0.07$; the dotted lines are 90% bootstrap confidence intervals. (ii) Number of quality bets for different values of $\tau$ out of the 760 betting opportunities in the two seasons.

The average returns in Figure 6(i) is not a smooth function of $\tau$. This is partly due to the role of longshots in this exercise. For example, at $\tau = 0.11$, we obtained 74 longshots from which 8 have been correct resulting in a net profit of 5.07 units. Even when we bet with 0.3 units for longshots, the betting strategy remains highly variable because for another value of $\tau$, another small number of correct longshots is obtained that can lead to a very different net profit. A more advanced betting strategy takes into account the variation of odds amongst the bookmakers. We abstain from such more advanced strategies since we only want to illustrate the performance of our model in a basic and simple betting strategy. The presented results can be used as a benchmark for the more advanced betting strategies based on our model. No definitive conclusions can be drawn from this illustration. We regard this validation study as 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 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 attack and defense strengths 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. Our empirical study is for a dataset of match results from nine seasons of the English Premier League. The last two seasons are 2010/11 and 2011/12 and 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 believe that we have presented promising results, we also believe that further improvements can be made in different directions. First, other dynamic model specifications for the attack and defense strengths can be considered such as random walk or long memory processes. Also we can include specific effects for summer and winter breaks in a sequence of football seasons. Second, our statistical modelling framework only uses 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 traveling distance of the visiting team. Third, as football betting in the United Kingdom is very popular, large sums of money are wagered on matches in the English Premier League. One can expect that odds provided by bookmakers are highly efficient. The odds used in our forecasting study are averages of odds provided by 28 to 40 bookmakers, depending on the match. In the 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.

# APPENDICES

## A Upper bound for correlation coefficient

Assume that $X$ and $Y$ are from the bivariate Poisson distribution with means $\lambda_x + \gamma$ and $\lambda_y + \gamma$, respectively, where $\gamma = \rho\sqrt{m_x m_y}$ where $m_x = \lambda_x + \gamma$ and $m_y = \lambda_y + \gamma$; see the definitions in Section 2.1. Since $\lambda_x, \lambda_y \geq 0$, we have $m_x \geq \gamma$ and hence $\rho \leq \sqrt{m_y/m_x}$. Similarly, we have $m_y \geq \gamma$ and $\rho \leq \sqrt{m_x/m_y}$. The upper bound for $\rho$ is given by

$$\rho \leq \min\left\{\sqrt{\frac{\lambda_x+\gamma}{\lambda_y+\gamma}}, \sqrt{\frac{\lambda_y+\gamma}{\lambda_x+\gamma}}\right\}.$$

# B Simulated maximum likelihood estimation

## B.1 Likelihood evaluation

Given our model specification for the time series of pairs of counts collected in $y$ with its dependence on the states in $z$, we can express the likelihood function $\ell(\psi)$ as given by (12). The individual observations and states at time $t$ are indicated by $y_t$ and $z_t$, respectively; see the discussion in Section 2.4. We evaluate the integral numerically by the method of importance sampling as developed by Shephard and Pitt (1997) and Durbin and Koopman (1997), hereafter referred to as SPDK. A comprehensive treatment of the method, together with other and related methods, is provided by Durbin and Koopman (2012, Part II). The SPDK method is based on an approximating linear Gaussian model $\mathrm{g}(y, z; \psi)$ which allows us to compute the approximate likelihood function $\mathrm{g}(y; \psi)$ by means of the Kalman filter and to simulate random samples for $z$ from $\mathrm{g}(z|y; \psi)$ by means of the simulation smoother; see the discussions in Jungbacker and Koopman (2007). The simulated random samples for $z$ will give a better support to $y$ although they come from an approximating model.

The likelihood function of the approximating Gaussian model $\mathrm{g}(y, z; \psi) = \mathrm{g}(y; \psi)\mathrm{g}(z|y; \psi)$ can be expressed as

$$\ell_{\mathrm{g}}(\psi) = \mathrm{g}(y; \psi) = \frac{\mathrm{g}(y, z; \psi)}{\mathrm{g}(z|y; \psi)} = \frac{\mathrm{g}(y|z; \psi)p(z; \psi)}{\mathrm{g}(z|y; \psi)}, \tag{15}$$

since $p(z; \psi) \equiv \mathrm{g}(z; \psi)$. Substituting $p(z; \psi) = \mathrm{g}(y; \psi)\mathrm{g}(z|y; \psi)/\mathrm{g}(y|z; \psi)$ into (12), we obtain

$$\ell(\psi) = \mathrm{g}(y; \psi) \int \frac{p(y|z; \psi)}{\mathrm{g}(y|z; \psi)} \mathrm{g}(z|y; \psi)\mathrm{d}z = \ell_{\mathrm{g}}(\psi)\mathbb{E}_{\mathrm{g}} \left\{ \frac{p(y|z; \psi)}{\mathrm{g}(y|z; \psi)} \right\}, \tag{16}$$

where $\mathbb{E}_{\mathrm{g}}$ refers to expectation with respect to the Gaussian density $\mathrm{g}(z|y; \psi)$. This method has proved to work effectively for multivariate time series models; see, for example, Koopman and Lucas (2008). In our model specification, the individual observations $y_t$ are independent for given $z_t$ as implied by (8) for $t = 1, \ldots, n$. Hence we can also assume that $\mathrm{g}(y|z; \psi) = \prod_{t=1}^{n} \mathrm{g}(y_t|z_t; \psi)$. The construction of an approximating model is discussed in Section B.2.

For a given approximating model, we estimate the likelihood function via Monte Carlo simulation as

$$\hat{\ell}(\psi) = \ell_{\mathrm{g}}(\psi) \frac{1}{M} \sum w_i, \qquad w_i = \frac{p(y|z^i; \psi)}{\mathrm{g}(y|z^i; \psi)}, \qquad z^i \sim \mathrm{g}(z|y; \psi), \tag{17}$$

where $w_i$ is referred to as an importance weight, $\ell_{\mathrm{g}}(\psi)$ is obtained from the Kalman filter and $z^i$ is computed by the simulation smoother for $i = 1, \ldots, M$. We can refer to $\hat{\ell}(\psi)$ as the importance sampling estimate of the likelihood function. For the purpose of likelihood maximisation with respect to $\psi$, it is preferred to work with the loglikelihood function. Taking the log of $\hat{\ell}(\psi)$ in (17) introduces a bias that can be accounted for in the usual way; see Durbin and Koopman (1997).

The effectiveness of the importance sampling method for likelihood evaluation relies on the properties of the importance sampling weight function $w(y, z; \psi) = p(y|z^i; \psi) \,/\, \mathrm{g}(y|z^i; \psi)$; see Geweke (1989) who provides conditions for $w(y, z; \psi)$ under which a central limit theorem is valid for the estimate $\hat{\ell}(\psi)$. An important condition is the existence of a variance for weight function $w(y, z; \psi)$. Based on a sample of importance weights $w_1, \ldots, w_M$, Koopman, Shephard, and Creal (2009) discuss diagnostic test statistics to validate the existence of a variance for the importance sampling weights.

## B.2 Construction of approximating model

For the implementation of the SPDK importance sampling method, the approximating linear Gaussian state space model is given by

$$\mathrm{g}(y, z; \psi) = \mathrm{g}(y|z; \psi)\mathrm{g}(z; \psi) = \mathrm{g}(z; \psi) \prod_{t=1}^{n} \mathrm{g}(y_t|z_t; \psi), \tag{18}$$

where $\mathrm{g}(z; \psi)$ represents the density of the dynamic state process (6) and we let $\mathrm{g}(y_t|z_t; \psi)$ be represented by the linear Gaussian model equation

$$y_t = a_t\delta + A_t z_t + c_t + \varepsilon_t, \qquad \varepsilon_t \sim \mathrm{NID}(0, V_t), \qquad t = 1, \ldots, n, \tag{19}$$

or more explicitly

$$\mathrm{g}(y_t|z_t; \psi) = \mathrm{NID}(a_t\delta + A_t z_t + c_t, V_t), \qquad t = 1, \ldots, n, \tag{20}$$

where vector $a_t$ has element 1 if the number of goals in the corresponding element of $y_t$ is from a home team and 0 otherwise, matrix $A_t$, with elements of 1s, 0s and -1s, selects the attack (+1) and defense (-1) strengths of the relevant teams, and mean correction $c_t$ and variance $V_t$ are selected such that the first and second derivates of logdensities $\log p(y_t|z_t; \psi)$ and $\log \mathrm{g}(y_t|z_t; \psi)$ with respect to $z_t$ are equal to each other, for $t = 1, \ldots, n$. We notice that $a_t\delta + A_t z_t$ represents the signal as also defined in (9). Closed-form solutions of these two sets of $n$ equalities are not available and hence we solve them iteratively with the use of the Kalman filter and smoother; more details and discussions are given by Jungbacker and Koopman (2007). The approximating model $\mathrm{g}(y, z; \psi)$ is effectively a second-order Taylor expansion of the true model and it is also equivalent to computing the mode of $p(z|y; \psi)$ for $z$; see the discussions in Durbin and Koopman (1997), So (2003) and Jungbacker and Koopman (2007). Our application for the bivariate Poisson model is not straightforward and we require to provide some further clarification. We will briefly discuss these necessary details for a successful implementation next.

To obtain values for $c_t$ and $V_t$ in (19), we need to solve the equations

$$\dot{\mathrm{g}}_t(z_t) = \dot{p}_t(z_t), \qquad \ddot{\mathrm{g}}_t(z_t) = \ddot{p}_t(z_t), \qquad t = 1, \ldots, n,$$

where

$$\dot{p}_t(z_t) = \frac{\partial \log p(y_t|z_t;\psi)}{\partial z_t}, \qquad \ddot{p}_t(z_t) = \frac{\partial^2 \log p(y_t|z_t;\psi)}{\partial z_t \partial z_t'},$$

and $\dot{\mathrm{g}}_t(z_t)$ and $\ddot{\mathrm{g}}_t(z_t)$ are defined similarly. It follows straightforwardly that

$$\dot{\mathrm{g}}_t(z_t) \equiv A_t' V_t^{-1}(y_t - c_t - a_t\delta - A_t z_t), \qquad \ddot{\mathrm{g}}_t(z_t) \equiv -A_t' V_t^{-1} A_t, \qquad t = 1, \ldots, n.$$

The derivatives for $\log p(y_t|z_t;\psi)$ are more intricate and we develop expressions for $\dot{p}_t(z_t)$ and $\ddot{p}_t(z_t)$ in the next section. Hence we obtain expressions for $c_t$ and $V_t$ by

$$V_t = -A_t \ddot{p}_t^{-1}(z_t) A_t', \qquad c_t = y_t - a_t\delta - A_t\left[z_t + \ddot{p}_t^{-1}(z_t)\dot{p}_t(z_t)\right], \qquad t = 1, \ldots, n. \tag{21}$$

We notice that matrix $A_t^{-1}$ exists in our framework. The mean $c_t$ and variance $V_t$ depend on the state vector $z_t$ and hence we solve these equations iteratively. For starting values of $c_t$ and $V_t$, we construct the linear Gaussian state space model for $\mathrm{g}(y, z;\psi)$ and apply the Kalman filter smoother to obtain $\hat{z} = \mathbb{E}_{\mathrm{g}}(z|y;\psi)$. From the value $z = \hat{z}$, we can obtain new values for $c_t$ and $V_t$ and can construct or update a new approximating model. The Kalman filter smoother produces a new $\hat{z}$ and we iterate this process until convergence. When this process has converged, the linear Gaussian model with the final values for $c_t$ and $V_t$ represents the approximating model $\mathrm{g}(y, z;\psi)$ as given by (19). It is well established that the Kalman filter and related methods can treat missing observations straightforwardly; see the discussions in Durbin and Koopman (2012, Part I).

## B.3 The derivatives for the model observation density

Equation (8) implies that the matches played at time $t$, for a given $z_t$, are treated as independent events. Hence we can treat each match separately. A match is for home team $i$ and visiting team $j$. The scoring intensities for both teams are collected in the $2 \times 1$ vector $\lambda_{ijt} = (\lambda_{x,ijt}, \lambda_{y,ijt})'$ which are functions of $z_t$, that is $\lambda_{ijt} = s_{ij}(z_t)$ since $\lambda_{x,ijt} = s_{x,ij}(z_t)$ and $\lambda_{y,ijt} = s_{y,ij}(z_t)$; see the discussion in Section 2.4. The first derivative of the log of the bivariate Poisson density (1) with respect to $z_t$ can be obtained via the chain rule as

$$\frac{\partial \log p(X, Y; \lambda_{x,ijt}, \lambda_{y,ijt};\gamma)}{\partial z_t} = \dot{s}_{ij}(z_t) \times \dot{p}_\lambda(\lambda_{ijt}),$$

where $X$ and $Y$ are specific elements of $y_t$ and represent the numbers of goals scored by teams $i$ and $j$, respectively, at time $t$, and where

$$\dot{s}_{ij}(z_t) = \frac{\partial \lambda_{ijt}'}{\partial z_t}, \qquad \dot{p}_\lambda(\lambda_{ijt}) = \frac{\partial \log p(X, Y; \lambda_{x,ijt}, \lambda_{y,ijt};\gamma)}{\partial \lambda_{ijt}}.$$

The second derivative can be obtained in the same way, that is

$$\frac{\partial^2 \log p(X, Y; \lambda_{x,ijt}, \lambda_{y,ijt};\gamma)}{\partial z_t \partial z_t'} = \dot{s}_{ij}(z_t) \times \ddot{p}_\lambda(\lambda_{ijt}) \times \dot{s}_{ij}(z_t)',$$

where

$$\ddot{p}_\lambda(\lambda_{ijt}) = \frac{\partial^2 \log p(X, Y; \lambda_{x,ijt}, \lambda_{y,ijt}; \gamma)}{\partial \lambda_{ijt} \partial \lambda'_{ijt}}.$$

An expression for $\dot{s}_{ij}(z_t)$ is obtained easily for link functions $s_{x,ij}(z_t)$ and $s_{y,ij}(z_t)$ as given by (3).

The general expressions for $\dot{p}_\lambda(\lambda_{ijt})$ and $\ddot{p}_\lambda(\lambda_{ijt})$ follow from (1) and are decomposed as

$$\dot{p}_\lambda(\lambda_{ijt}) = \begin{pmatrix} \dot{p}_{\lambda_x}(\lambda_{ijt}) \\ \dot{p}_{\lambda_y}(\lambda_{ijt}) \end{pmatrix}, \qquad \ddot{p}_\lambda(\lambda_{ijt}) = \begin{bmatrix} \ddot{p}_{\lambda_{xx}}(\lambda_{ijt}) & \ddot{p}_{\lambda_{xy}}(\lambda_{ijt}) \\ \ddot{p}_{\lambda_{xy}}(\lambda_{ijt}) & \ddot{p}_{\lambda_{yy}}(\lambda_{ijt}) \end{bmatrix}. \tag{22}$$

The first derivative elements are given by

$$\dot{p}_{\lambda_x}(\lambda_{ijt}) = \lambda_{x,ijt}^{-1}[X - \lambda_{x,ijt} - U(1, \lambda_{ijt})], \quad \dot{p}_{\lambda_y}(\lambda_{ijt}) = \lambda_{y,ijt}^{-1}[Y - \lambda_{y,ijt} - U(1, \lambda_{ijt})],$$

where $U(m, \lambda) = S(m, \lambda)/S(0, \lambda)$ with

$$S(m, \lambda) = \sum_{k=0}^{min(X,Y)} \binom{X}{k} \binom{Y}{k} k!\, k^m \left(\frac{\gamma}{\lambda_x\, \lambda_y}\right)^k,$$

and with $\lambda = (\lambda_x, \lambda_y)'$ for $m = 0, 1, 2$. We notice that

$$\frac{\partial S(m, \lambda)}{\partial \lambda_u} = -\lambda_u^{-1} S(m + 1, \lambda), \qquad u = x, y, \qquad m = 0, 1,$$

and $S(m, \lambda) = 0$ when $\gamma = 0$, for $m = 1, 2$. We further observe that $S(0, 0) = 1$ so that function $U(m, \lambda)$ is properly defined for all $\gamma \geq 0$. The second derivative elements are given by

$$\begin{aligned} \ddot{p}_{\lambda_{xx}}(\lambda_{ijt}) &= -\lambda_{x,ijt}^{-1} \left[1 + \dot{p}_{\lambda_x}(\lambda_{ijt}) - \lambda_{x,ijt}^{-1} \dot{U}(\lambda_{ijt})\right], \\ \ddot{p}_{\lambda_{yy}}(\lambda_{ijt}) &= -\lambda_{y,ijt}^{-1} \left[1 + \dot{p}_{\lambda_y}(\lambda_{ijt}) - \lambda_{y,ijt}^{-1} \dot{U}(\lambda_{ijt})\right], \\ \ddot{p}_{\lambda_{xy}}(\lambda_{ijt}) &= \lambda_{x,ijt}^{-1} \lambda_{y,ijt}^{-1} \dot{U}(\lambda_{ijt}), \end{aligned}$$

with

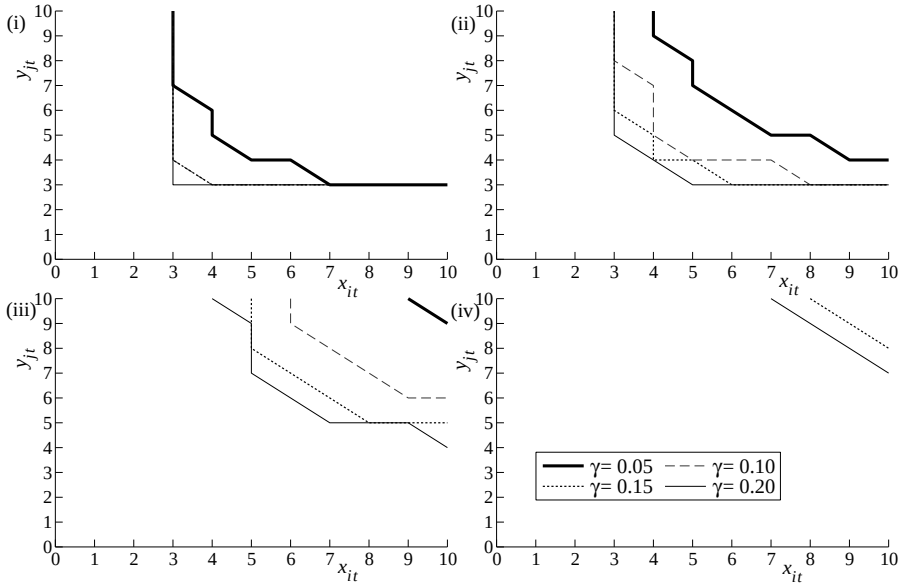
$$\dot{U}(\lambda) = U(2, \lambda) - U(1, \lambda)^2, \qquad \frac{\partial U(1, \lambda)}{\partial \lambda_u} = -\lambda_u^{-1} \dot{U}(\lambda), \qquad u = x, y.$$

Finally, it follows that

$$\dot{p}_t(z_t) = \sum_{i,j \in y_t} \dot{s}_{ij}(z_t) \times \dot{p}_\lambda(\lambda_{ijt}), \qquad \ddot{p}_t(z_t) = \sum_{i,j \in y_t} \dot{s}_{ij}(z_t) \times \ddot{p}_\lambda(\lambda_{ijt}) \times \dot{s}_{ij}(z_t)',$$

where the notation $i, j \in y_t$ implies that we consider all matches played at time $t$ with a home team $i$ and a visiting team $j$, for $t = 1, \ldots, n$.

FIGURE 7: POSITIVE, NEGATIVE, AND INDEFINITE AREAS OF THE HESSIAN MATRIX




The figure illustrates combinations of counts which generate positive, negative and indefinite "variances" in the approximating model, for different values of $\lambda_x$, $\lambda_y$ and $\gamma$. The areas below and left from the lines correspond to counts that generate positive variances. The areas above and right from the lines represent counts that provide negative or indefinite variances. The coefficient $\gamma$ ranges from 0.05 to 0.20 with 0.05 increments. The panels are for (i) $\lambda_x = \lambda_y = 1.0$; (ii) $\lambda_x = 1.5, \lambda_y = 1.0$; (iii) $\lambda_x = 2.0, \lambda_y = 1.5$; (iv) $\lambda_x = 2.5, \lambda_y = 2.0$.

## B.4 Computational issues

The construction of the approximating model and the generation of the importance samples require the application of the Kalman filter smoother applied to the linear Gaussian model (19). Since matrix $V_t$ in (21) is a variance matrix, we require that $V_t$ is positive definite or that $\ddot{p}_t^{-1}(z_t)$ is negative definite which effectively insists that the $2 \times 2$ matrix $\ddot{p}_\lambda(\lambda)$ in (22) is negative definite. Jungbacker and Koopman (2007) have argued that even when $V_t$ is not positive definite, the application of the Kalman filter and the corresponding computations are still appropriate for our purposes. However, it is insightful to verify under which conditions $\ddot{p}_\lambda(\lambda)$ in (22) is negative. We therefore need to verify the determinant of $\ddot{p}_\lambda(\lambda)$. Without providing the details, we present in Figure 7 the values of $X$ and $Y$ for which we obtain a positive definite matrix $\ddot{p}_\lambda(\lambda)$. In case $\gamma = 0$, the variance $V_t$ is well defined since the model reduces to a double Poisson which imposes a proper variance; see Durbin and Koopman (2012, Chapter 10.6) for the details. In case $\gamma > 0$, the variance $V_t$ becomes negative when $X$ and/or $Y$ are large in relation to their intensities $\lambda_x$ and/or $\lambda_y$, respectively. The benchmark values can be deduced from Figure 7.

# References


Barnett, V. and S. Hilditch (1993). The effect of an artificial pitch surface on home team performance in football (soccer). *J. Royal Statistical Society A 156(1)*, 39–50.

Crowder, M., M. J. Dixon, A. Ledford, and M. Robinson (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.

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

Diebold, F. X. and R. S. Mariano (1995). Comparing predictive accuracy. *Journal of Business and Economic Statistics 13*, 253–265.

Dixon, M. J. and S. G. Coles (1997). Modelling association football scores and inefficiencies in the football betting market. *Applied Statistics 46(2)*, 265–280.

Durbin, J. and S. J. Koopman (1997). Monte Carlo maximum likelihood estimation for non-Gaussian state space models. *Biometrika 84(3)*, 669–684.

Durbin, J. and S. J. Koopman (2012). *Time Series Analysis by State Space Methods* (2nd ed.). Oxford: Oxford University Press.

Geweke, J. (1989). Bayesian inference in econometric models using Monte Carlo integration. *Econometrica 57*, 1317–39.

Giacomini, R. and H. White (2006). Tests of conditional predictive ability. *Econometrica 74(6)*, 1545–1578.

Goddard, J. (2005). Regression models for forecasting goals and match results in association football. *International Journal of Forecasting 21*, 331–340.

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

Jungbacker, B. and S. J. Koopman (2007). Monte Carlo estimation for nonlinear non-Gaussian state space models. *Biometrika 94(4)*, 827–839.

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

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

Koopman, S. J. and A. Lucas (2008). A non-Gaussian panel time series model for estimating and decomposing default risk. *J. Business and Economic Statist. 26*, 510–25.

Koopman, S. J., N. Shephard, and D. D. Creal (2009). Testing the assumptions behind importance sampling. *J. Econometrics 149*, 2–11.

Koopman, S. J., N. Shephard, and J. A. Doornik (2008). *SsfPack 3.0: Statistical algorithms for models in state space form*. London: Timberlake Consultants Press.

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

Ord, K., C. Fernandes, and A. C. Harvey (1993). Time series models for multivariate series of count data. In T. S. Rao (Ed.), *Developments in Time Series Analysis*. London: Chapman and Hall.

Rue, H. and O. Salvesen (2000). Prediction and retrospective analysis of soccer matches in a league. *Journal of the Royal Statistical Society: Series D (The Statistician) 49(3)*, 399–418.

Shephard, N. and M. K. Pitt (1997). Likelihood analysis of non-Gaussian measurement time series. *Biometrika 84(3)*, 653–667.

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

So, M. K. P. (2003). Posterior mode estimation for nonlinear and non-Gaussian state space models. *Statistica Sinica 13*, 255–274.