TI 2017-062/III
Tinbergen Institute Discussion Paper

Forecasting Football Match Results in National League Competitions Using Score-Driven Time Series Models

Siem Jan (S.J.) Koopman[1]
Rutger Lit[2]

1: VU Amsterdam, The Netherlands; CREATES, Aarhus University, Denmark; Tinbergen Institute, The Netherlands
2: VU Amsterdam, The Netherlands

Forecasting football match results in national league competitions using score-driven time series models

Siem Jan Koopman[a,b] *and Rutger Lit*[a]

[a]Vrije Universiteit Amsterdam and Tinbergen Institute, The Netherlands

[b]CREATES, Aarhus University, Denmark

July 3, 2017

Abstract

We develop a new dynamic multivariate model for the analysis and the forecasting of football match results in national league competitions. The proposed dynamic model is based on the score of the predictive observation mass function for a high-dimensional panel of weekly match results. Our main interest is to forecast whether the match result is a win, a loss or a draw for each team. To deliver such forecasts, the dynamic model can be based on three different dependent variables: the pairwise count of the number of goals, the difference between the number of goals, or the category of the match result (win, loss, draw). The different dependent variables require different distributional assumptions. Furthermore, different dynamic model specifications can be considered for generating the forecasts. We empirically investigate which dependent variable and which dynamic model specification yield the best forecasting results. In an extensive forecasting study, we consider match results from six large European football competitions and we validate the precision of the forecasts for a period of seven years for each competition. We conclude that our preferred dynamic model for pairwise counts delivers the most precise forecasts and outperforms benchmark and other competing models.

Key words: Football, Forecasting, Score-driven models, Bivariate Poisson, Skellam, Ordered probit, Probabilistic loss function.

1 Introduction

Forecasting football match results is a highly popular activity. Amongst football supporters, it is widespread to make a forecast of the next match result and oftentimes the individual forecast is positively biased towards the team that one is supporting. But even the pundit knows that forecasting a match result is a challenging task. The common way to bet on a football match is simply to indicate whether one expects the team to win, lose, or draw its next game. Whether the match result is a win, loss or draw depends on the difference in the number of goals scored by the two opposing teams in a football match. There are many determining factors of scoring a goal including the attack strength of the team, the defence strength of the opposing team, the home ground advantage (when applicable), and specific events taking place during the match. We consider the use of three possible observational variables to base our forecast of the next match result in terms of win, loss, or draw. The first variable is two-dimensional and consists of the number of goals scored by the two opposing teams during a match. The second variable is the difference between the number of goals scored. The third variable is simply the indicator of win, loss, or draw. The informational content of these three consecutive variables is clearly decreasing. For each of the variable categories, a variety of dynamic models can be considered for the forecasting of the match result. Many contributions in the statistical literature on the modelling and forecasting of the three variables have been made. We refer to Table B.1 in the Appendix for a schematic overview of the main contributions. A discussion of this earlier literature is next and is followed by a discussion of our contributions to this literature.

Most contributions in the statistical literature on the modelling and forecasting of match results focus on the first variable where the pairwise observations of numbers of goals scored by the opposing teams are assumed to come from a bivariate distribution. The probability for a possible match outcome is implied by the bivariate distribution and is formally given by $P(X = x, Y = y)$, for $x, y \in \mathbb{N}_0$, where X and Y denote the number of goals scored by the home and away team, respectively. Hence we have a probability for any match outcome. The main interest usually focuses on forecasting the probabilities of home win, draw, or away win; these are the *toto probabilities* and are given by $P(X > Y)$, $P(X = Y)$, and $P(X < Y)$, respectively. The parameters of the distribution can be expressed as function of strengths of attack and defence of the competing teams. This procedure was first proposed by Maher (1982) who expresses the means of the double-Poisson distribution (product of two independent Poissons) as team-specific strengths of attack and defence. Dixon and Coles (1997) consider the double-Poisson distribution as well and introduce a dependence parameter for the match results 0–0, 1–0, 0–1 and 1–1. They also propose a weighting function to down-weight likelihood contributions of observations from the more distant past. In Crowder, Dixon, Ledford, and Robinson (2002), the model of Dixon and Coles (1997) is formulated as a non-Gaussian state space model with time-varying strengths of attack and defence. Then they develop approximating methods for parameter estimation and signal extraction as they stated that an exact analysis is computationally too expensive. A bivariate Poisson distribution is also used by Karlis and Ntzoufras (2003) who show that the introduction of a parameter for dependence between goals scored by both teams during a match, leads to a more accurate prediction of the outcome of a draw. Rue and Salvesen (2000) incorporate the framework of Dixon and Coles (1997) and develop a dynamic generalized linear model which is analysed by Markov chain Monte Carlo methods in continuous time. Goddard (2005) explores the inclusion of covariates in a bivariate Poisson model. Koopman and Lit (2014) show that a high-dimensional panel of weekly match results can be analysed effectively within a non-Gaussian state space framework based on the bivariate Poisson model with stochastically time-varying attack and defence strengths, and with some of the above extensions. Their analysis includes the exact maximum likelihood estimation of the parameter vector and the exact signal extraction of the time-varying attack and defence strengths of the two teams. Detailed evidence of its forecast precision in forecasting match results is presented. Finally, another interesting and original contribution in this category is given by Dixon and Robinson (1998) who treat the number of scored goals by the competing teams during a match as interacting birth processes.

The second category is the difference between goals in a match and can be regarded as the margin of victory of a team. In this category we let $Z = X - Y$ be the difference between the number of scored goals X and Y, with $Z \in \mathbb{Z}$. By modelling Z, we consider the toto probabilities as given by $P(Z > 0)$, $P(Z = 0)$, and $P(Z < 0)$ for a home win, draw, and away win, respectively. By modelling the difference of goals, information is lost since, for example, the pairs $(X = 0, Y = 1)$ and $(X = 2, Y = 3)$ produce the same values for Z. On the other hand, a smaller number of summations is needed to obtain toto probabilities from Z when compared to the pair (X, Y). It is not immediately clear what the overall effect of modelling Z instead of (X, Y) would be on the forecasting of the toto probabilities. The reasoning behind this is the accumulation of modelling error which could potentially be smaller since a smaller number of probability components are summed compared to the first category. A model for the difference between goals in football matches is provided by Karlis and Ntzoufras (2009) who introduce the Skellam distribution for analysing match results. This distribution was originally derived by Skellam (1946) as the difference of two independent Poisson distributions. However, Karlis and Ntzoufras (2009) show that independence is not strictly necessary and even the Poisson assumption for the pair of variables (X, Y) is not needed. In their analysis the parameters of the Skellam distribution are kept static. Lit (2016, Ch. 4) extends the Skellam model to allow for strengths of attack and defence that evolve stochastically over time in a non-Gaussian state space framework.

Instead of modelling toto probabilities via the double or bivariate Poisson models or via the Skellam models, we can also consider the modelling of the toto probabilities directly. For this third variable category, we introduce ordered logit or ordered probit models in our study. The modelling of match results in terms of win, loss, and draw, rather than scores or differences in scores, leads to a more parsimonious model but also to a simpler estimation procedure. Koning (2000) investigates the balance in competition in Dutch professional soccer by means of an ordered probit model with static team strengths. A selection of covariates can be introduced in the static ordered probit regression model of Goddard and Asimakopoulos (2004) and in the static ordered logit model of Forrest and Simmons (2000). Cattelan, Varin, and Firth (2013) propose a (semi)-dynamic Bradley-Terry model in which team strengths are modelled by exponentially weighted moving average processes. An early contribution is made by Fahrmeir and Tutz (1994) who introduce an ordered logit non-Gaussian state space model that incorporates random walks for the team strengths. The estimation of parameters for this model is carried out by the Kalman filter and recursive posterior mode estimation methods. The dynamic cumulative link model of Knorr-Held (2000) has been applied to German Bundesliga data for an analysis based on the extended Kalman filter and smoother. Finally, Hvattum and Arntzen (2010) propose an ordered logit model in which team strengths are updated over time using a so-called Elo rating system.

Our research contributes to the literature in a number of ways. First, we develop a new dynamic multivariate model for the analysis and forecasting of football match results for each of the three variable categories. The dynamic extensions of the static models are based on the class of score-driven models where the time-varying coefficients are updated as an autoregressive process. The autoregressive updating of the time-varying parameter is driven by the score of the conditional observation probability density function, see Creal, Koopman, and Lucas (2013) for a discussion of this approach. Three features of this class of models are particularly attractive in our context: (i) The score-driven models are observation driven which means that the likelihood is available in closed form. This allows for a fast estimation process despite the high-dimensional model challenges due to the large number of teams that participate in an European football competition over a number of years. The computationally more demanding Kalman filter is not required for estimation and forecasting; (ii) The filtered estimates of the time-varying parameters in a score-driven model are locally optimal in a Kullback-Leibler sense, see Blasques, Koopman, and Lucas (2015); (iii) The forecasting performance of the score-driven models is comparable to their parameter-driven counterparts, see Koopman, Lucas, and Scharth (2016). Second, we determine which of the three variable categories leads to the most accurate forecasts in an extensive empirical study. Third, we also investigate, as part of our empirical study, whether dynamic models with time-varying parameters show better forecasting performance when compared to models with static parameters. We further verify whether the dynamic extension of the static model is best achieved by formulating a time-varying parameter model or by weighting the likelihood contributions over time as proposed in Dixon and Coles (1997).

We have constructed time series panels of match results from six European competitions: the English Premier League, the German Bundesliga, the Spanish Primera División, the French Ligue 1, the Italian Serie A, and the Dutch Eredivisie. We have collected 17 seasons of match results ranging from 1999–2000 to 2015–2016, for which the first 10 seasons are used for parameter estimations and the last 7 seasons for the forecasting study. The size of our forecasting study allows us to draw strong conclusions with respect to the forecasting performances of the considered models. We use the rank probability score as a loss function and explain why this is the most suitable loss function for this exercise. The losses are evaluated by the Diebold and Mariano (1995) statistic to test for equal predictive accuracy.

The remainder of this research report is organized as follows. We introduce the statistical modelling framework in Section 2 where the specific details of the score-driven football models are discussed in Section 3. We discuss the design of our extended forecasting study, including a data description, present our empirical findings and discuss various aspects of our analyses in Section 4. Section 5 concludes. An Appendix provides additional figures, tables, and technical details including the score functions of the various distributions that are considered.

2 The distributions for the three variable categories

For the three variable categories, we develop three different modelling frameworks. First we consider the different observational characteristics and propose their corresponding discrete mass functions. Their dynamic extensions are developed and discussed in Section 3.

2.1 Bivariate Poisson distribution

The outcome of a football match is simply determined by the number of goals scored and conceded by a team. The outcome can be considered as a pair of counts (X, Y) where X is the number of goals scored by the home team and Y by the away team. We may assume that the pair of counts (X, Y) is generated by a bivariate Poisson distribution with intensities $\lambda_1, \lambda_2 > 0$ for (X, Y) and with the covariance between (X, Y) denoted by $\lambda_3 \geq 0$. The probability mass function of the bivariate Poisson distribution is given by

$$p_{\mathrm{BP}}(X = x, Y = y; \lambda_1, \lambda_2, \lambda_3) = e^{-(\lambda_1+\lambda_2+\lambda_3)} \frac{\lambda_1^x}{x!} \frac{\lambda_2^y}{y!} \sum_{k=0}^{min(x,y)} \binom{x}{k} \binom{y}{k} k! \left(\frac{\lambda_3}{\lambda_1 \lambda_2} \right)^k, \qquad (1)$$

see Kocherlakota and Kocherlakota (1992) and Johnson, Kotz, and Balakrishnan (1997) for more information. It can be shown that

$$\mathbb{E}_{\mathrm{BP}}(X) = \mathbb{V}\mathrm{ar}_{\mathrm{BP}}(X) = \lambda_1 + \lambda_3, \quad \mathbb{E}_{\mathrm{BP}}(Y) = \mathbb{V}\mathrm{ar}_{\mathrm{BP}}(Y) = \lambda_2 + \lambda_3, \quad \mathbb{C}\mathrm{ov}_{\mathrm{BP}}(X, Y) = \lambda_3, \quad (2)$$

where $\mathbb{E}_p$, $\mathbb{V}\mathrm{ar}_p$ and $\mathbb{C}\mathrm{ov}_p$ denotes expectation, variance and covariance, respectively, with respect to density p. For $\lambda_3 = 0$ the bivariate Poisson distribution reduces to the double Poisson distribution. The covariance is a "shared component" in the intensities: a higher λ_3 leads typically to a higher number of equal observations ($X = Y$) which for a football match is a draw.

In the context of modelling match results in football, we follow the framework developed by Maher (1982); it has become the standard in the statistics literature on sports modelling. We therefore specify the intensities λ_1 and λ_2 as functions of the latent strengths of attack α and defence β of the two opposing teams, and the home ground advantage effect δ. Suppose home team i welcomes away team j for a football match. Then intensity $\lambda_{1,ij}$, associated with the number of home goals X in this match of team i versus team j, and intensity $\lambda_{2,ij}$, associated with the corresponding number of away goals Y, can be specified as

$$\lambda_{1,ij} = \exp(\delta + \alpha_i - \beta_j), \qquad \lambda_{2,ij} = \exp(\alpha_j - \beta_i), \qquad (3)$$

with α_m and β_m being the attack and defence strengths, respectively, of team $m = i, j$ and $i \neq j$. The home ground advantage δ can also be made team-specific but we restrict this effect to be equal for all teams.

Assume we have a data set of football match results in a competition of N teams, for a number of yearly football seasons, and which have taken place in a total of T weekends (and mid-weeks) or football rounds. The data vectors with match results for round t are ordered consecutively over time, $t = 1, \ldots, T$. Hence we have a time series panel. The outcome of a match between home team i and away team j is recorded by the number of home goals x_{ijt} and the number of away goals y_{ijt}, for $i, j = 1, \ldots, N$, with $i \neq j$, and $t = 1, \ldots, T$.

We notice that for each round t in the competition, we have $N/2$ matches with $N/2$ home teams and $N/2$ away teams. For this data set and with the assumption that the match result (X, Y) is independently generated by the bivariate Poisson model described above, we can specify the log-likelihood function as

$$\mathcal{L}_{\mathrm{BP}}(\psi_{\mathrm{BP}}) = \sum_{t=1}^{T} \sum_{i=1}^{N/2} \log p_{\mathrm{BP}}(x_{ijt}, y_{ijt}; \lambda_{1,ij}, \lambda_{2,ij}, \lambda_3), \qquad i = i(t), \quad j = j(i,t),$$

where i is a function of t, since it represents the home team in round t, and where j is a function of i and t, since it represents the opponent of the home team i in round t, that is $i = i(t)$ and $j = j(i, t)$. Given the specifications of the intensities in (3), the parameter vector ψ_{BP} consists of attack and defence strengths α_m and β_m, respectively, for $m = 1, \ldots, N$, together with the home ground advantage δ and the static covariance λ_3. Hence the number of unknown coefficients is $2(N + 1)$. The maximum likelihood estimate of ψ_{BP} is obtained via the numerical maximisation of the log-likelihood function $\mathcal{L}_{\mathrm{BP}}(\psi_{\mathrm{BP}})$ with respect to ψ_{BP}. The maximisation typically relies on gradient-based methods such as the Newton-Raphson method. Although the parameter vector is typically of a high-dimension, the gradients (score function and information matrix) have closed-form expressions. For further details of exact maximum likelihood estimation and alternative estimation methods for the bivariate Poisson distribution, we refer to Holgate (1964), Gourieroux, Monfort, and Trognon (1984), Karlis and Ntzoufras (2003), and Kocherlakota and Kocherlakota (2001).

Once the parameter vector ψ_{BP} is estimated for the bivariate Poisson distribution, the probabilities of a win, draw, and loss for the home team i against the away team j in a football match are given by

$$\begin{aligned} P(X_{ij} > Y_{ij}) &= \sum_{x=1}^{\infty} \sum_{0 \le y < x} p_{\mathrm{BP}}(x, y; \hat{\lambda}_{1,ij}, \hat{\lambda}_{2,ij}, \hat{\lambda}_3), \\ P(X_{ij} = Y_{ij}) &= \sum_{x=y=0}^{\infty} p_{\mathrm{BP}}(x, y; \hat{\lambda}_{1,ij}, \hat{\lambda}_{2,ij}, \hat{\lambda}_3), \\ P(X_{ij} < Y_{ij}) &= \sum_{x=0}^{\infty} \sum_{y > x} p_{\mathrm{BP}}(x, y; \hat{\lambda}_{1,ij}, \hat{\lambda}_{2,ij}, \hat{\lambda}_3), \end{aligned} \tag{4}$$

respectively, where $\hat{\lambda}_{k,ij}$ and $\hat{\lambda}_3$ are the maximum likelihood estimates of coefficients $\lambda_{k,ij}$ and λ_3, respectively, for $k = 1, 2$. The estimate $\hat{\lambda}_3$ is directly obtained from the maximum likelihood estimate of ψ_{BP} while the estimates $\hat{\lambda}_{k,ij}$, for $k = 1, 2$, are constructed from it using (3). In practice, the infinite upper bound is replaced by 25 which gives sufficient accuracy.

2.2 Skellam distribution

The win, loss or draw of a football match is determined by the difference between the numbers of goals scored and conceded by a team. The difference between the numbers of home goals X and away goals Y can be regarded as the margin of the victory of a team. We can assume this difference of the counts $Z = X - Y$ to be distributed by the Skellam distribution with intensities λ_4, λ_5. The probability mass function of the Skellam distribution is given by

$$p_{\mathrm{Sk}}(Z = z; \lambda_4, \lambda_5) = e^{-(\lambda_4+\lambda_5)} \, (\lambda_4/\lambda_5)^{z/2} \, I_{|z|}(2\sqrt{\lambda_4\lambda_5}), \tag{5}$$

where $I_{|z|}(\cdot)$ is the modified Bessel function of order $|z|$. The mean and variance of Z are given by

$$\mathbb{E}_{\mathrm{Sk}}(Z) = \lambda_4 - \lambda_5, \qquad \mathbb{V}\mathrm{ar}_{\mathrm{Sk}}(Z) = \lambda_4 + \lambda_5. \tag{6}$$

We refer to the original work of Skellam (1946) and Irwin (1937) for the derivation of the Skellam distribution based on the difference of two independent Poisson distributions. Alzaid and Omair (2010) presented higher moments and several other interesting properties of the Skellam distribution. Karlis and Ntzoufras (2009) showed that the underlying Poisson assumption is not strictly necessary and that the Skellam distribution can also be considered by itself as a distribution defined on integers. The Skellam distribution can alternatively be formulated directly in terms of location and scale parameters which enables the modelling of the mean and variance explicitly; see Koopman, Lit, and Lucas (2017) who adopted this formulation to extract stochastic volatility from discrete price changes in financial markets.

When modelling football match results in terms of their victory margins, we can also incorporate the framework of Maher (1982) for the Skellam distribution. The intensity λ_4 is associated with the number of home goals and intensity λ_5 with the number of away goals. Hence the specifications for $\lambda_{1,ij}$ and $\lambda_{2,ij}$ in (3) can apply similarly to $\lambda_{4,ij}$ and $\lambda_{5,ij}$, respectively, for home team i and away team j. Assume we have a similar data set as described above but now we only record the goal difference for a match of home team i versus away team j in round t, this is $z_{ijt} = x_{ijt} - y_{ijt}$. For this data set and with the assumption that the margin of victory Z of a match is independently generated by the Skellam model as described above, we can specify the log-likelihood function as

$$\mathcal{L}_{\mathrm{Sk}}(\psi_{\mathrm{Sk}}) = \sum_{t=1}^{T} \sum_{i=1}^{N/2} \log p_{\mathrm{Sk}}(z_{ijt}; \lambda_{4,ij}, \lambda_{5,ij}), \qquad i = i(t), \quad j = j(i,t),$$

where the functions $i(t)$ and $j(i,t)$ are described above. The parameter vector ψ_{Sk} is of dimension $2N+1$ which is one less than ψ_{BP} since λ_3 is not present in the Skellam distribution. When estimating ψ_{Sk} the same issues arise for its maximum likelihood estimation as with estimating ψ_{BP}.

Once the maximum likelihood estimate of ψ_{Sk} for the Skellam model is obtained, the probabilities of a win, draw, and loss for the football match of home team i against away team j are given by

$$\begin{aligned} P(Z_{ij} > 0) &= \sum_{z=1}^{\infty} p_{\text{Sk}}(z; \hat{\lambda}_{4,ij}, \hat{\lambda}_{5,ij}), \\ P(Z_{ij} = 0) &= p_{\text{Sk}}(0; \hat{\lambda}_{4,ij}, \hat{\lambda}_{5,ij}), \\ P(Z_{ij} < 0) &= \sum_{z=-\infty}^{-1} p_{\text{Sk}}(z; \hat{\lambda}_{4,ij}, \hat{\lambda}_{5,ij}), \end{aligned} \tag{7}$$

respectively, where $\hat{\lambda}_{k,ij}$ is the maximum likelihood estimate of $\lambda_{k,ij}$, for $k = 4, 5$, and can be constructed from equation (3) and the maximum likelihood estimate of ψ_{Sk}. In practice, the infinite upper bound is replaced by 25 which gives sufficient accuracy.

2.3 Ordered probit models

The win, loss or draw of a football match can also be considered as an observed variable that we then model directly. In this case the observed categorical variable C is simply determined by $C = 2$ for a home win $X > Y$ (or $Z > 0$), $C = 1$ for a draw $X = Y$ (or $Z = 0$), and $C = 0$ for a home loss $X < Y$ (or $Z < 0$). The margin of victory is not measured. The variable C can also be interpreted as the credit points for a win, draw or loss of a match, although in all our considered football competitions, the credit for a win is 3 points rather than 2. In an ordered probit model, we assume that an unobserved stochastic variable C^* determines the category C probabilistically, with C^* given by the equation

$$C^* = \lambda_6 + \eta, \qquad \eta \sim \mathcal{N}(0, \sigma_\eta^2), \tag{8}$$

where λ_6 is an unknown constant that indicates the strength of the home team relative to the away team, and where η is a random variable generated by a normal variable with mean zero and variance σ_η^2. For the ordered probit model, we assume that the variable C is generated conditional on C^* via the equations

$$C = \begin{cases} 2 & \text{if} \quad C^* \leq \kappa_1, \\ 1 & \text{if} \quad \kappa_1 < C^* \leq \kappa_2, \\ 0 & \text{if} \quad C^* > \kappa_2, \end{cases} \tag{9}$$

where the cutoff points κ_1 and κ_2, and the variance σ_η^2 are treated as unknown parameters; see, for example, Greene (2012) for a textbook treatment of ordered probit models and also for a more general treatment with more categories . Given the construction with the unobserved variable C^* and the random variable η, we cannot jointly identify the three parameters uniquely. Therefore, we constrain the scale of η and set $\sigma_\eta^2 = 1$. An alternative is to set κ_1 to zero and choose σ_η^2 freely, together with κ_2. For an ordered probit model with categorical observations $C \in \{2, 1, 0\}$ and $\sigma_\eta^2 = 1$, the probability density function is given by

$$p_{\mathrm{OP}}(C \in \{2,1,0\}; \lambda_6, \kappa_1, \kappa_2) = \begin{cases} \Phi(\kappa_1 - \lambda_6) & \text{if } \; C = 2, \\ \Phi(\kappa_2 - \lambda_6) - \Phi(\kappa_1 - \lambda_6) & \text{if } \; C = 1, \\ 1 - \Phi(\kappa_2 - \lambda_6) & \text{if } \; C = 0, \end{cases} \tag{10}$$

where $\Phi(\cdot)$ is the standard normal cumulative density function (cdf). To ensure that the probabilities are all positive, we further restrict the parameters by $\kappa_1 < \kappa_2$.

Given the limited amount of information in the category variable of a win, loss or draw of a football match, the framework of Maher (1982) cannot be incorporated in an ordered probit model. It is also the design of the ordered probit model that does not allow the separation of the strength of a team in attack and defence strengths. Since the relative strength of the home team is represented by λ_6 and since it determines the probability of category $C \in \{2, 1, 0\}$, we have

$$\lambda_{6,ij} = \gamma_i - \gamma_j, \tag{11}$$

where γ_m is the total strength or capability of team m.

Assume we have a similar data panel of match results as described above but now with only a record of the match result as a win, loss or draw, for home team i versus away team j in round t, that is $c_{ijt} \in \{2, 1, 0\}$. For this data set and with the assumption that the category variable $C \in \{0, 1, 2\}$ is independently generated by the ordered probit model as described above, we can specify the log-likelihood function as

$$\mathcal{L}_{\mathrm{OP}}(\psi_{\mathrm{OP}}) = \sum_{t=1}^{T} \sum_{i=1}^{N/2} \log p_{\mathrm{OP}}(c_{ijt}; \lambda_{6,ij}, \kappa_1, \kappa_2), \qquad i = i(t), \quad j = j(i,t),$$

where the functions $i(t)$ and $j(i,t)$ are described above. Given the specification (11), the parameter vector ψ_{OP} consists of strengths γ_m, for $m = 1, \ldots, N$, together with the cutoff constants κ_1 and κ_2. Hence the number of unknown coefficients is $N + 2$. The maximum likelihood estimation can be carried out by a gradient-based optimisation method applied to $\mathcal{L}_{\mathrm{OP}}(\psi_{\mathrm{OP}})$ with respect to ψ_{OP}; analytical expressions are available for the gradients and facilitate fast computation. To let $\kappa_1 < \kappa_2$, the coefficient κ_2 from parameter vector ψ_{OP} can be replaced by κ_2^* with $\kappa_2 = \kappa_1 + \exp(\kappa_2^*)$. We notice that the home ground advantage δ is implicitly accounted for by the cutoff parameters κ_1 and κ_2 because they are uniquely associated with the probabilities of home and away wins, respectively. In effect, the difference between their absolute values is the home ground advantage.

Given the maximum likelihood estimate of ψ_{OP} for the ordered probit model, we can evaluate the probabilities of a win, draw, and loss of home team i against away team j which are given by

$$\begin{aligned}
P(C_{ij} = 2) &= \Phi(\hat{\kappa}_1 - \hat{\lambda}_{6,ij}), \\
P(C_{ij} = 1) &= \Phi(\hat{\kappa}_2 - \hat{\lambda}_{6,ij}) - \Phi(\hat{\kappa}_1 - \hat{\lambda}_{6,ij}), \\
P(C_{ij} = 0) &= 1 - \Phi(\hat{\kappa}_2 - \hat{\lambda}_{6,ij}),
\end{aligned} \tag{12}$$

respectively, where $\Phi(\cdot)$ is the standard normal cdf, $\hat{\lambda}_{6,ij}$ and $\hat{\kappa}_k$ are the maximum likelihood estimates of $\lambda_{6,ij}$ and κ_k, respectively, for $k = 1, 2$; the estimates $\hat{\lambda}_{6,ij}$ can be constructed from (11) and the maximum likelihood estimate of ψ_{OP}.

3 Score-driven time-varying parameters

The treatment of football match results using the three observational variables discussed in Section 2 is relatively straightforward. This is partly due to the assumptions that matches and the efforts by the teams in each round of the competition are treated as independent events. However, it is not realistic to assume that the numbers of goals scored by a team in a series of matches are treated as an independent events. The strength of a football team is likely to be partly related to the performance of the team in recent matches. The attack and defence strengths of teams change also over time when the compositions of teams evolve through the years. Several dynamic extensions of the static models described above are considered in the statistics and econometrics literature. We consider a selection of such existing methods in Section 3.4 for comparisons.

We contribute however by developing an effective and computationally fast approach to the dynamic modelling of attack and defence strengths of football teams. These developments are presented in Section 3.1 where a short review is given of score-driven time series models and in Section 3.2 where the details of its implementation are presented for the models of Section 2. The initialization of the dynamic processes are discussed in Section 3.3. In Section 4 we present the empirical results from European football league competitions; they include excellent forecasting results for our proposed dynamic extensions.

3.1 Score-driven time series models: a short review

We consider the class of score-driven models of Creal et al. (2013) to capture the dynamic behavior of a parameter or a selection of parameters. We treat the case of a panel of time series variables for which y_t represents the corresponding vector of observations at time t. We assume that the data are generated from a distribution with density function $p\left(y_t; \psi\right)$ for which the density functions discussed in Section 2 are examples. The observation vector y_t can include the pairs of counts for the number of goals of both teams, the differences between the number of goals in a match, or the indicator of a win, loss or draw. A part of the static model parameters in ψ is assumed to be time-varying and is collected in the time-varying parameter vector f_t. The remaining static parameters are collected in the parameter vector ψ^*.

In this framework, the score-driven model is based on the predictive density function that is treated as the observation density and is given by

$$y_t \sim p\left(y_t | f_t, \mathcal{F}_t; \psi^*\right), \qquad t = 1, \ldots, T, \tag{13}$$

where $\mathcal{F}_t$ represents the information set available at time t, consisting of lagged observations $\{y_{t-1}, y_{t-2}, \ldots\}$ and past time-varying parameter vectors $\{f_{t-1}, f_{t-2}, \ldots\}$. The score-driven updating mechanism for the time-varying parameter f_t is given by

$$f_{t+1} = \omega + B f_t + A s_t, \tag{14}$$

where ω is a vector of unknown constants, matrices A and B are unknown coefficient matrices, and s_t is the scaled score vector as defined by

$$s_t = S_t \cdot \nabla_t, \qquad \nabla_t = \frac{\partial \log p(y_t | f_t, \mathcal{F}_t; \psi^*)}{\partial f_t}, \qquad S_t = S(f_t, \mathcal{F}_t; \psi^*), \tag{15}$$

with $S(\cdot)$ being a matrix function to scale the score vector. A score-driven model updates the factor f_{t+1} in the direction of the steepest increase of the log-density at time t given the current parameter f_t and the data history $\mathcal{F}_t$. Under correct model specification, the score vectors are a martingale sequence since $\mathbb{E}_{t-1}(s_t) = 0$ where $\mathbb{E}_{t-1}$ denotes the expectation with respect to $p(y_t | f_t, \mathcal{F}_t; \psi^*)$.

The scaling matrix is regularly chosen to be a function of the variance of the score to take into account the curvature of the log-density at time t as summarized by the Fisher information matrix $\mathcal{I}_{t|t-1} = \mathbb{E}_{t-1}\left[\nabla_t \nabla_t'\right]$. When it is intricate or impossible to obtain the Fisher matrix analytically, we can take S_t as the unity matrix. The score-driven updating of parameters has a theoretical foundation since the estimates of the time-varying parameter are optimal in a Kullback-Leibler sense, see Blasques et al. (2015), and it is therefore not a heuristic method.

A nice feature of the score-driven time series model is the availability of the likelihood function in closed form. In particular, the log-likelihood function is given by

$$\mathcal{L}(\psi) = \sum_{t=\tau}^{T} \log p\left(y_t | f_t, \mathcal{F}_t; \psi^*\right),$$

for some given parameter value for ψ^*, where integer $\tau > 0$ is set for initialisation purposes, see the discussion below. The time series observations y_t are all given, for $t = 1, \ldots, T$, and the time-varying parameter values f_t are evaluated recursively as in (14), for some initial value f_1 and for a given value ψ^*. These simple and fast computations also lead to a much faster optimization of the likelihood compared to the parameter driven counterparts. In particular, when compared to the simulation-based methods that are required for the log-likelihood evaluation for non-Gaussian state space models; see Section 3.4 for further discussions. The aspect of simple and fast computations becomes especially important when we consider the high-dimensional time series panels that we have in mind for the modelling of match results in football competitions. Finally, Koopman et al. (2016) have shown that the forecasting performance of score-driven models is similar or highly competitive to their parameter-driven counterparts, including the state space formulations. Their study however has only considered univariate models whereas in our study we investigate the forecasting performance of multivariate score-driven models.

The predictive density $p\left(y_t | f_t, \mathcal{F}_t; \psi^*\right)$ is conditional on the time-varying parameter at time t and the data history $\mathcal{F}_t \in \{y_{t-1}, y_{t-2}, \ldots, f_{t-1}, f_{t-2}, \ldots\}$. We notice the reliance of the recursion (14) on past data since the scaled score s_t is clearly a function of y_t. Hence f_t is a function of $\{y_{t-1}, y_{t-2}, \ldots\}$. The dynamic extensions of the models in Section 2 are achieved simply by substituting a static parameter by a time-varying parameter. There is no direct involvement of lagged y_t's in the density functions of Section 2 and hence we can drop $\mathcal{F}_t$ from the conditional set to obtain $p\left(y_t | f_t; \psi^*\right)$. This is still a predictive density and it represents one of the densities in Section 2, that is $p\left(\cdot | f_t; \psi^*\right)$ has the same functional form as $p_{\mathcal{M}}(\cdot; \psi_{\mathcal{M}})$ with $\mathcal{M} \in \{\mathrm{BP}, \mathrm{Sk}, \mathrm{OP}\}$.

3.2 Score-driven models for football match results

Next we adopt the score-driven time-varying parameter framework for the three densities discussed in Section 2. We consider the time-variation for a selection of parameters in ψ

and provide details for its implementation for modelling football match results. We obtain a flexible and effective framework for the time series analysis and forecasting of football match results in large competitions and over many seasons of a competition.

Bivariate Poisson distribution

The modelling of football match results via the observation pair (x_{ijt}, y_{ijt}), where x_{ijt} and y_{ijt} are the numbers of goals by the home team i and the away team j, respectively, in round t, can be based on the bivariate Poisson distribution $p_{\mathrm{BP}}(X = x_{ijt}, Y = y_{ijt}; \lambda_{1,ij}, \lambda_{2,ij}, \lambda_3)$ and using the approach of Maher (1982) as reflected in the specifications (3). The dynamic model allows us to let the strengths of the teams in attack and defence be time-varying. In particular, we replace α_i and β_i in (3) by α_{it} and β_{it}, respectively. We then obtain the $2N \times 1$ time-varying parameter f_t which contains α_{it} and β_{it} for all N teams active in a competition, that is

$$f_t = (\alpha_{1t}, \ldots, \alpha_{Nt}, \beta_{1t}, \ldots, \beta_{Nt})', \qquad t = 1, \ldots, T. \tag{16}$$

The home ground advantage δ and the covariance λ_3 in ψ_{BP} can remain constant over time: they can be treated as static parameters and are placed in ψ^* of the score-driven model. The implication of this dynamic extension is that the intensities $\lambda_{k,ij}$ can now be treated as time-varying intensities that we denote by $\lambda_{k,ijt} = \lambda_{k,ij}(f_t)$, for $k = 1, 2$, where $\lambda_{k,ij}()$ refers to the functions in (3).

The time-varying updating equation for f_t is provided by (14). However, it is more efficient to carry out the updating at round t for each match result. We assume that the observation pair (x_{it}, y_{jt}) is generated by the bivariate Poisson and we select a subset of f_t that is relevant for this match, that is

$$f_{ijt} = (\alpha_{it}, \alpha_{jt}, \beta_{it}, \beta_{jt})' = M_{ij} f_t, \tag{17}$$

where M_{ij} is the $4 \times 2N$ selection matrix of 0s and 1s, and is implicitly defined. We update the selected time-varying parameters as in (14). It reduces to the updating

$$f_{ij,t+1} = \omega_{ij} + B_{ij} f_{ijt} + A_{ij} s_{ijt}, \tag{18}$$

with 4×1 vector of constants $\omega_{ij} = M_{ij}\omega$, 4×4 coefficient matrices $A_{ij} = M_{ij} A\, M_{ij}'$ and $B_{ij} = M_{ij} B\, M_{ij}'$, and with the 4×1 scaled score vector s_{ijt} which is defined as (15) but with the gradient with respect to 4×1 vector f_{ijt}. This updating is then repeated for all

$N/2$ matches in each round t and hence the full vector f_t is effectively updated. Given the update f_{t+1}, we can make predictions for the match results for the next round $t+1$ in the competition. In particular, we can forecast the probabilities of a win, draw, and loss as described at the end of Section 2.1.

The dimension of vector f_t is as high as $2N$. We may therefore want to specify the $2N \times 2N$ coefficient matrices A and B in a parsimonious manner. In our empirical study, we specify them as

$$A = \begin{bmatrix} a_1 \cdot I_N & 0 \\ 0 & a_2 \cdot I_N \end{bmatrix}, \qquad B = \begin{bmatrix} b_1 \cdot I_N & 0 \\ 0 & b_2 \cdot I_N \end{bmatrix},$$

such that the attack strengths (α) rely on updating coefficients a_1 and b_1 and the defence strengths (β) rely on updating coefficients a_2 and b_2. The coefficient matrices A_{ij} and B_{ij} are then defined similarly but for $N = 2$. In this specification, the coefficient matrices A and B do not have team specific characteristics and there are no spillover effects in the dynamic specifications between the teams. Such specifications can be considered in a straightforward manner because all unknown elements of A and B are placed in ψ^*. Finally, the specification of the scaled score $s_{ijt} = S_{ijt} \cdot \nabla_{ijt}$ in (18) for the bivariate Poisson is provided in Appendix A.2, at least for the derivation of ∇_{ijt}. Since the derivation of the Fisher matrix is intricate, we set the scaling to the unity matrix, that is $S_{ijt} = I_4$. The parameter vector ψ^* is given by

$$\psi^* = (a_1, a_2, b_1, b_2, \lambda_3, \delta)', \tag{19}$$

and is estimated by maximum likelihood. The estimation of ω_{ij} is discussed in section 3.3.

A team plays only once in each round t. Hence the maximum number of matches at time t is $N/2$. When every team plays according to schedule, a season consists of $T = 2(N-1)$ rounds in a competition. In practice, however, some football matches are postponed due to bad weather conditions and other external events. These matches are then played later at a convenient time. For this purpose, additional rounds are inserted in the calendar of the competition. A small set of matches are scheduled for such additional rounds. If a team does not play in round t, its score is set to zero and the updating for the strengths of attack and defence reduces to

$$\alpha_{m,t+1} = \omega_m + b_1 \alpha_{mt}, \qquad \beta_{m,t+1} = \omega_m + b_2 \beta_{mt},$$

respectively, for any team m that does not play in round t.

Skellam distribution

When the football match result is modelled in terms of the observed margin of victory, $z_{ijt} = x_{ijt} - y_{ijt}$, we can opt for the Skellam distribution in Section 2.2 with a similar dynamic extension as for the bivariate Poisson distribution. The probability mass function of the Skellam distribution is given by (5); also see Koopman et al. (2017) for a discussion of its implementation in a dynamic setting. The time-varying parameter f_t has the same composition as in (16) and hence we consider the Skellam pmf $p_{\text{Sk}}(Z = z_{ijt}; \lambda_{4,ijt}, \lambda_{5,ijt})$ where $\lambda_{k,ijt} = \lambda_{k,ij}(f_t)$, for $k = 4, 5$, where $\lambda_{k,ij}()$ refers to the corresponding functions in (3). The updating of f_t can take place for each match separately using (18) where f_{ijt} is defined in (17). The derivation of the score s_{ijt} for the Skellam density is presented in Appendix A.3. We set the scaling equal to the unity matrix, that is $S_{ijt} = I_4$. The parameter vector ψ^* is given by $\psi^* = (a_1, a_2, b_1, b_2, \delta)'$ and is estimated by maximum likelihood.

Ordered probit

When we record the data simply by win, loss or draw of a football match, that is we observe $c_{ijt} \in \{2, 1, 0\}$, we can model the data by the ordered logit model of Section 2.3 with a dynamic extension for the overall strength (or capability) of the team. Given that we cannot separate the strength in defence and attack, we have a more parsimonious model. The strengths are made time-varying by replacing γ_i by γ_{it} and place them in the $N \times 1$ vector f_t. We have

$$f_t = (\gamma_{1t}, \ldots, \gamma_{Nt})', \qquad t = 1, \ldots, T.$$

As a result, the coefficient $\lambda_{6,ij}$ has also become time-varying, we define $\lambda_{6,ijt} = \gamma_{it} - \gamma_{jt}$ and it indicates the difference in strength between the home team i and away team j for their football match in round t. This is our dynamic version of the ordered logit model for football match results of Koning (2000). The cutoff points κ_1 and κ_2 (which also represent the home ground advantage) remain static coefficients.

The updating of the time-varying team capabilities f_t can also be done for each match result separately as implied by (18) but now f_{ijt} is simply the 2×1 vector $(\gamma_{it}, \gamma_{jt})'$, with 2×1 constant vector $\omega_{ij} = (\omega_i, \omega_j)'$ and 2×2 coefficient matrices $A_{ij} = a_1 \cdot I_2$ and $B_{ij} = b_1 \cdot I_2$. The derivation of the 2×1 score vector s_{ijt} for the ordered probit pmf is provided in Appendix A.4. We set the scaling equal to the unity matrix, that is $S_{ijt} = I_2$. The parameter vector is given by $\psi^* = \{a_1, b_1, \kappa_1, \kappa_2\}$ and is estimated by maximum likelihood.

3.3 Initialization of the panel

The updating equation for the time-varying parameter vector f_t requires an initial value f_1. Given the high-dimensional f_t for our high-dimensional panel of time series, this is not a straightforward task. Next we describe several ways of initializing the updating equation for f_t and we discuss their pros and cons. In our empirical study we consider the match results for a series of consecutive competitions. Since almost all football competitions have a promotion and relegation system, there is a considerable set of teams that do not play at the start of the data set and only become active at a later stage. The teams that are not present in the first competition, are excluded from the initialization at time $t = 1$ but will receive a separate treatment; see below.

To illustrate the challenge of initialization, consider Figures B.1 and B.2 which display the numbers of goals scored and conceded, respectively, in the German Bundesliga competitions from 2000-01 to 2015-16. In this Bundesliga panel, there is a considerable number of teams that do not play for all years of our sample and a several number of teams only play one season. We discuss our initialization method for the dynamic bivariate Poisson model. The initialization for the Skellam and ordered probit models can be done in a similar fashion.

For the teams that play in the first year of the data set (and possibly for many more years), we consider two strategies. First, the elements of the $2N \times 1$ vector f_1 are estimated as part of ψ^* and the constant vector ω can be set to the unconditional mean of the score driven update function; we have $\omega = f_1 \odot (\mathbf{1} - \text{diagonal(B)})$ where $\odot$ denotes point-wise multiplication and where $\mathbf{1}$ is a vector of ones. A clear disadvantage of including f_1 in ψ^* is the large increase of its dimension. For the bivariate Poisson model, we need to add $2N$ additional parameters which need to be estimated, in addition to the parameters in equation (19). For the relatively small length of the time series dimension, the estimation of f_1 increases the uncertainty in the parameter space while forecast precision may suffer from this. Our second initialization strategy takes out the data from the first year of competition and is used to obtain static estimates of the strengths of attack and defence; see, for example, Maher (1982) in which a regression method is used. Then, f_1 can be set equal to the static estimates of the strengths of attack and defence. The parameters in the static regression are not identified, but this can be elevated by restricting the strengths of attack to sum to zero, that is $\sum_{i=1}^{N} \alpha_i = 0$. Once we have an estimate for f_1, vector ω can be determined as suggested above. This solution comes at the cost of a shorter data sample of one year.

We also consider two different treatments for football teams that enter the panel at a later point in time due to promotion. First, in case of the second initialization method (that is regression with the zero sum restriction for the attack strengths), we can expect that the strength of a top team is typically larger than zero while the strength of a poor team is smaller than zero. We introduce index τ that indicates the first round of a new competition in which a new team enters the panel. The elements of f_τ that correspond to the new team are simply set to zero, the middle point between a top and a poor team, because we do not know much about this new team. This strategy appears reasonable. In our second treatment of newly promoted teams, the elements of f_τ are given values such that the average attack strength of all teams at time $\tau - 1$, denoted by $\bar{\alpha}_{\tau-1}$, equals $\bar{\alpha}_\tau$. Same procedure can be applied to defence strengths: at round τ, we choose the defence strengths of newly promoted teams such that $\bar{\beta}_{\tau-1} = \bar{\beta}_\tau$. Since relegated teams have typically a low strength, the newly promoted teams are initialized with low strengths as well. This can be an undesirable feature of this method.

Finally, when football teams relegate during one competition but get promoted in a future competition, we emphasize that the updating continues also for "missing" observations. When such teams re-enter the competition, their strengths have probably reverted to their long-term mean which is reasonable.

3.4 Other dynamic extensions

Parameter-driven state space model

An alternative dynamic extension of the three discrete models in Section 2 is obtained by formulating a non-Gaussian state space model where the observation density function is specified conditional on a stochastically time-varying vector $\tilde{f}_t$, consisting of attack and defence strengths which are treated as latent dynamic variables, that is

$$y_t \sim p(y_t \mid \tilde{f}_t; \psi^*), \qquad \tilde{f}_{t+1} = \omega + B\tilde{f}_t + A\eta_t, \qquad \eta_t \sim \mathcal{N}(0, I), \tag{20}$$

for $t = 1, \ldots, T$, where ψ^*, ω, A and B play similar roles as those for the score-driven model discussed in Section 3.1, but they can have different values and will lead to different forecast functions. The key difference is that f_t in (14) is a function of past observations $y_{t-1}, y_{t-2}, \ldots$, while $\tilde{f}_t$ in (20) is a stochastic unobserved dynamic process. We notice that $p(y_t \mid \tilde{f}_t; \psi^*)$ represents one of the densities $p_{\mathcal{M}}(\cdot; \psi_{\mathcal{M}})$ from Section 2, for $\mathcal{M} \in \{\mathrm{BP}, \mathrm{Sk}, \mathrm{OP}\}$.

The partially non-Gaussian state space model (20) has been treated by Koopman and Lit (2014) in the context of modelling football match results with an application to the English Premier League. In particular, they consider the pair of counts (X, Y) and assume it is generated by the bivariate Poisson distribution with probability density function (1). The intensities of the distribution are specified as in (3) with the strengths of attack (α) and defence (β) in $\tilde{f}_t$. For the evaluation of the log-likelihood function, we need to integrate out all latent dynamic variables numerically using efficient simulation methods. This treatment of non-Gaussian observations in the context of the state space model (20) can also be considered for the dynamic Skellam model with stochastically time-varying strengths; see the treatment in Lit (2016, Ch. 4). Fahrmeir and Tutz (1994) and Knorr-Held (2000) have adopted a similar state space model for analyzing categorical football match outcomes of win, draw and loss.

Weighting likelihood method

The dynamic extensions of our models aim to increase the role of past observations in the modelling and forecasting of match results. To achieve this without considering a dynamic extension of the static models in Section 2, we can allow recent observations to be more influential than observations in the more remote past by a direct weighting method, see Dixon and Coles (1997) in the context of football match results. This direct method defines the log-likelihood function as

$$\mathcal{L}^W_{\mathcal{M}}(\psi_{\mathcal{M}}) = \sum_{t=1}^{T} \phi(t) \sum_{i=1}^{N/2} \log p_{\mathcal{M}}(\cdot;\psi_{\mathcal{M}}), \tag{21}$$

for $\mathcal{M} \in \{\mathrm{BP}, \mathrm{Sk}, \mathrm{OP}\}$ and where $\phi(t)$ is a non-increasing weighting function of index t and where the density $p_{\mathcal{M}}()$ represents one of the densities in Section 2. A typical example of a weighting function is $\phi(t) = \exp(-\xi t)$ which allows data contributions from the more distant past to be down weighted in entering the log-likelihood function. Although this method is relatively simple by construction, the unknown parameter ξ cannot be estimated by maximum likelihood because $\mathcal{L}^W_{\mathcal{M}}(\psi_{\mathcal{M}}) \to 0$ as $\xi \to \infty$, for any data set. Instead, we can choose ξ such that, for example, the sum of squared prediction errors is minimized. The empirical evidence in Dixon and Coles (1997) suggests that more precise forecasts can be obtained using this approach, when compared to standard maximum likelihood estimation of the parameters in the static model.

4 Forecasting football match results in Europe

Our empirical study is a basic and straightforward exercise: we forecast all match results in the next round of a football competition, for all rounds in seven yearly competitions and for six European football competitions. In this design of our study we almost make 15,000 probabilistic forecasts for the football toto results which are the match results in terms of win, loss and draw for the home team. These probability forecasts are based on a particular model. We consider all three static models from Section 2 (using the three different variable categories) and their various dynamic extensions as discussed in Section 3. This forecasting study is of an exceptional magnitude and hence it allows us to draw strong conclusions concerning which model performs best in forecasting. The forecast precision measurements are based on the average rank probability score statistics.

4.1 Data description

We forecast the football toto results for six European football competitions: The English Premier League, German Bundesliga, Spanish Primera División, Italian Serie A, France Ligue 1, and the Dutch Eredivisie. For each competition, the total data set consist of 17 seasons of football match results. We have partitioned the data set into the in-sample seasons 1999-2009 which is used for initial parameter estimation, and the out-of-sample seasons 2009-2016 which is used for our forecasting study. After each football season, the poorest performing team(s) will be relegated and new teams will be promoted into the competition. Hence the total number of teams in the data set increases with every season since the relegated teams remain in the panel as they can re-appear in future seasons. The number of relegated teams differs per competition and per season. We refer to Table 1 for some descriptive statistics of the six football competitions. The data used in our empirical study can be found at `http://www.football-data.co.uk`.

4.2 Estimated strengths from score-driven model

To empirically illustrate our proposed score-driven model as a dynamic extension of our models for football match results, we present in Figure 1 the time-varying estimates of the attack, defence and total strengths of the two major rival teams in the Spanish Primera división: Barcelona and Real Madrid. We present the estimated strengths for the dynamic bivariate Poisson model; the implementation details for filtering and parameter estimation are discussed in Section (3.2). The graphs in Figure 1 are all based on the values of f_t as defined in (16) and recursively evaluated by (14) where ω, A and B are replaced by their maximum likelihood estimates. The estimated toto probabilities are computed as in (4) but with the underlying parameters replaced by their estimates. Since the strengths are time-varying (α and β are in f_t), these probability computations are done repeatedly, each time before a new football round starts.

The estimated strengths of attack and defence for Barcelona and Real Madrid, as obtained from the dynamic Bivariate Poisson model, reveal that in the 16 seasons from 2000 onwards,

Table 1

Descriptive statistics of six football competitions

The table reports in-sample and out-of-sample characteristics of the six football competitions that are considered in the forecasting study. The column '# Teams' denote the number of teams that are active in one season of the respective competition. For the Italian Serie A, the in-sample data set has 5×18 and 5×20 teams in a season. For the French Ligue 1, the in-sample data set has 3×18 and 7×20 teams in a season. Column titles Mean(·) and Var(·) denote the sample mean and sample variance of the home (H) and away (A) goals.

Competition	# Teams	# Matches	Mean(H)	Mean(A)	Var(H)	Var(A)
In-sample 1999-2009						
English Premier League	20	3800	1.503	1.092	1.649	1.181
German Bundesliga	18	3060	1.673	1.185	1.781	1.274
Spanish Primera división	20	3800	1.533	1.116	1.566	1.169
Italian Serie A	18/20	3430	1.505	1.089	1.434	1.109
French Ligue 1	18/20	3578	1.382	0.913	1.381	0.960
Dutch Eredivisie	18	3060	1.766	1.242	2.195	1.478
Out-of-sample 2009-2016						
English Premier League	20	2660	1.573	1.171	1.752	1.312
German Bundesliga	18	2142	1.617	1.281	1.804	1.462
Spanish Primera división	20	2660	1.627	1.121	1.947	1.343
Italian Serie A	20	2660	1.496	1.120	1.497	1.174
French Ligue 1	20	2660	1.420	1.050	1.409	1.151
Dutch Eredivisie	18	2142	1.789	1.325	1.956	1.558

the attack strengths have been competitive and steadily increasing for both teams while the defence strength of Barcelona has been overall stronger since 2004 and has become even more stronger in the more recent years. The overall superior strength of Barcelona over Real Madrid since 2004 has been small but nevertheless clearly visible. This conclusion is also supported by the probabilities of a Barcelona win compared to a Real Madrid win. However, here we find that, since 2008 rather than 2004, the probability for a Barcelona win is persistently close to 0.5 while the Real Madrid win probability is closer to 0.35 during the last 7 seasons of the sample. We notice that the strengths and the probabilities are displayed for each match in the sample, hence the strengths of both teams are not exclusively presented for the Barcelona against Real Madrid (and vice versa) matches. To focus more on those two matches in each season, we have indicated in each plot when such a key match took place and whether it was a win, a draw or a loss for Barcelona. During the 16 seasons in our sample, the rivals have played 32 times against each other: 14 wins for Barcelona, 8 draws and 10 wins for Real Madrid. The home ground advantage effect is not accounted for in these plots in order to have more precise comparisons.

Similar graphs as in Figure 1 can be presented for the dynamic extension of the Skellam model while only the total strengths can be presented for the dynamic ordered probit model since the separation into attack and defence strengths cannot be identified in the latter framework. The total strengths and the toto probability estimates for the dynamic extensions of the Skellam and ordered probit models are presented in Figure B.3 of Appendix B. When we compare these results amongst the three score-driven models, the paths of the estimated total strengths are clearly different although the main patterns appear to be similar. Hence the question of which model is best in forecasting the toto outcome is relevant and of interest. We have presented the results for the two rival teams from the Spanish Primera división, but, for completeness, we present similar figures to Figure 1 with panels of attack, defence and overall strengths and a panel of toto probabilities, for the two main rival teams in the other national football league competitions; see Figures B.4-B.8 in Appendix B. These results merely illustrate that the estimated attack, defence and overall strengths are truly time-varying. It is interesting to view the strength increases in the last years of our sample by teams such as Dortmund, Juventus, and Paris SG but also the more recent strength decreases of Manchester United.

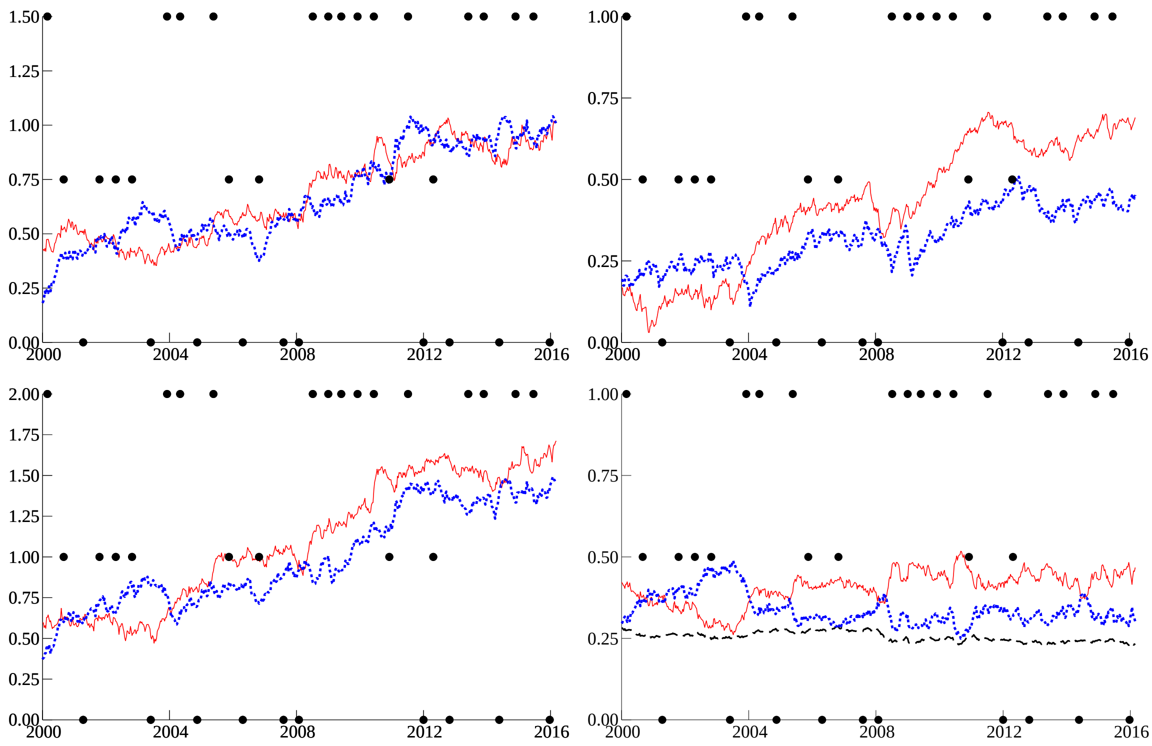
4.3 Forecasting: design of study and precision measurement

We produce probability forecasts for the toto outcomes of the next round of matches in six national league competitions and based on nine model categories as described in the introduction and summarised in Table B.1 of Appendix B. For a description of the static models we refer to, for example, Maher (1982), Karlis and Ntzoufras (2003), and Koning (2000) and for a description of the semi-dynamic models we refer to Dixon and Coles (1997). The details of our dynamic extension based on score-driven models are discussed in Section 3. The probability forecasts for the toto results are computed as implied by (4), (7), and (12), where the strengths of attack and defence (or overall) are either treated as static or as time-varying.

Before we compute the forecasts for round $t + 1$, all static parameters (whether in ψ or in ψ^*) are re-estimated using all data up to time t. The first forecasts are for the toto probabilities of all matches in the first round of the football season 2009-2010 and are based on the parameter estimates from the data panel of the previous ten seasons 1999-2009. These computations are repeated for each model, dynamic extension and method. In case of the score driven model, we recursively evaluate f_t and at the end of the estimation sample we obtain f_{t+1} from which, together with the static parameter estimates, the probability forecasts can be computed. Given the realised match results and their forecasts, we can evaluate a loss function to measure the forecast precision; see the details below. For the next round of football matches and its forecasts, we re-estimate the parameter vector after

Figure 1

Estimated strengths for Barcelona and Real Madrid



The estimated attack, defence and total strengths from the score-driven dynamic extension of the bivariate Poisson model for the two top teams in the Spanish Primera división: Barcelona (solid/red) and Real Madrid (dotted/blue). All panels: a dot on top of the panel represents a win for Barcelona, dot at the bottom is a win for Real Madrid, and a draw is represented by a dot in the middle of the panel. Top left panel: time series plot of extracted strengths of attack. Top right panel: time series plot of extracted strengths of defence. Bottom left panel: sum of extracted strengths of attack and defence. Bottom right panel: probability of toto results from the dynamic Bivariate Poisson model: Barcelona win (solid/red), Real Madrid win (dotted/blue) and draw (dashed/black). In these graphs, the home ground advantage is not taken into account, that is δ is set to zero in (3), to be able to better compare the two teams between each other.

including the football match results of the most recent round in our data set. Hence we have an expanding estimation sample, to ensure that we can utilize as much data for estimation. The procedure for our forecasting study is therefore simple, after each round of matches: re-estimation of the static parameters, filtering of the time-varying parameters (if any), and forecasting of the toto probabilities in the next round. We repeat these steps for each round in the seven consecutive football seasons and for each of the six European football competitions.

Given the forecasted probabilities for a win, loss and draw of the match and the realised toto result, for each match in a round, we can measure the precision of our forecasts for this round as follows. For example, assume that we have two rival models that produce probability forecasts for the toto outcome of a football match: Model I has $P(\text{win}) = 0.50$,

$P(\text{draw}) = 0.40$, and $P(\text{loss}) = 0.10$; Model II has $P(\text{win}) = 0.50$, $P(\text{draw}) = 0.30$, and $P(\text{loss}) = 0.20$. The outcome of the match is a home win. If we opt for the log-loss function, we would have $\log(0.50)$ as the loss for this match for both models, while the assigned probabilities to the other possible outcomes $P(\text{draw})$ and $P(\text{away})$ are ignored. The rank probability score (RPS) is a loss function that accounts for the other probabilities as well; see Epstein (1969) and, for an application to football match results, see Constantinou and Fenton (2012). For the toto forecasts, the RPS statistic is given by

$$RPS = 0.5 \sum_{k=1}^{3} \left(cdf_{f,k} - cdf_{y,k} \right)^2 . \tag{22}$$

where $cdf_{f,k}$ and $cdf_{y,k}$ are the cumulative density functions of the forecast and realised outcome, respectively. In our example, we have $RPS_I = [(0.5-1)^2+(0.9-1)^2+(1-1)^2]/2 = 0.13$ and $RPS_{II} = [(0.5-1)^2+(0.8-1)^2+(1-1)^2]/2 = .145$ for Models I and II, respectively. Hence the probability forecasts from Model I have been more precise. We average the RPS statistic over all football matches in a round and we take this average as our loss function. To facilitate model comparisons, we collect the value of this loss function, for each round, in a loss vector and use it to compute the Diebold Mariano (DM) test statistic for equal predictive accuracy; see Diebold and Mariano (1995). The DM statistic is asymptotically distributed as a standard normal random variable and hence rejects the null hypothesis of equal predictive accuracy at the 5% level of significance if the DM test statistic is smaller than −1.96 (the benchmark model performs significantly worse) or larger than 1.96 (the benchmark model performs significantly better). Finally, we also report the ARPS which is defined as the average of the RPS statistic over all rounds in the football season, and over all seven years in our out-of-sample data set (it is simply the average of the values in the loss vector).

4.4 Results of forecasting study

In Tables 2 and 3 we present the ARPS and the DM statistics for our three static models, their dynamic extensions and different initialization methods. We report these results for the six European football competitions and represent a summary of our findings. Given that for each competition we have made ±2,500 forecasts on average, we may regard our forecasting study as impressive. From the reported results, we learn that our dynamic extension, based on the score-driven model with f_1 being estimated using the first season of the in-sample data set, is the best performing forecasting strategy for all six European football competitions. It is only for the Spanish Primera división that a constrained version of our model with

$b1 = b_2 = 1$ (leading to a random walk updating for the time-varying parameter f_t) is the best performing in forecasting.

Our score-driven dynamic extensions outperform the static model and the semi-dynamic model extension in forecast precision significantly, in almost all cases. The Dutch Eredivisie is the exception: the outperformance is not strongly significant. We also learn that estimating f_1 as part of ψ^* brings too much uncertainty in the parameter space as the associated forecasts are significantly worse almost in all cases. The best strategy for initialization, that is setting a value for f_1, appears to be obtained by estimating the static model using the first season in the sample.

Tables 2 and 3 provide convincing evidence that the bivariate Poisson is the preferred distribution with the Skellam distribution in second place. These two models are almost always preferred when compared to the ordered probit model. The Dutch Eredivisie is the only exception where the dynamic ordered probit model is preferred in terms of forecast precision. However the superiority of the ordered probit model is never significant when compared to the other dynamic models. We may conclude that the condensation of data has a negative impact on forecast precision: we loose information when data is recorded in a more condensed manner. The counts of number of goals for both teams in a football match contain more information than the difference of these two counts, and even much more so than the sign of the difference (and zero). However, we could also have opted for another bivariate distribution than the Poisson. For example, we could have opted for the bivariate negative Binomial distribution; see Famoye (2010) for all relevant details. Its dynamic extension can be implemented in a similar way as for the Poisson but the scaled score function for updating the time-varying parameters will be different. However, the reported data descriptives in Table 1 do not give much evidence of over-dispersion in the number of goals scored, perhaps there is only some evidence for the Dutch Eredivisie competition. Although we do not report these results, we have produced the forecasting results for the bivariate negative Binomial model but we have not found any improvements when compared to the Poisson model.

Koopman et al. (2016) have argued that the forecasting performance of univariate score-driven models is comparable to their parameter-driven (state space model) counterparts. In our study we have confirmed this conclusion but now for a class of multivariate score-driven and state space models. In terms of forecast precision, the score-driven model produces a lower forecast loss than the dynamic state space model; in some instances we even report a significant improvement. Finally, we also report the number of seconds of computer-time needed for maximizing the log-likelihood function for a single model. The differences in computing-time for parameter estimation is noteworthy: estimation requires < 10 seconds for the score-driven model and approximately one hour for the state space model. The score-driven models clearly outperform the state space model, both in terms of forecast precision and computer-time.

5 Conclusion

We have developed a multivariate score-driven model to analyse a high-dimensional panel of football match results. The score-driven methodology is applied to three classes of models. In the first class, a match result is treated as a pairwise observation which is assumed to come from the bivariate Poisson distribution. The second class of models assumes that the difference between the number of goals, or the margin of victory of a team, is generated by the Skellam distribution. In the third class of models, the possibility of a win, draw, or loss of a match is modelled by an ordered probit model. These different model classes with their different variables require somewhat different statistical treatments but they can be extended with time-varying parameters using the same score-driven framework. All three approaches are able to forecast toto probabilities for football matches in a national league competition. In a large-scale forecasting study we have investigated which of the three model classes performs best in forecasting the toto probabilities in the next round of the competition. For this purpose, we have used a large panel match results from six European football competitions over a range of seasons. The results of the forecasting study show that our score-driven football models outperform a range of benchmark models in forecast precision but also in computing time. The dynamic bivariate Poisson model turns out to be the best performing model in forecasting overall while the ordered probit model does almost never produce a more precise forecast. We may conclude that the subsequent merging of data (from two counts, to the difference in counts, on to the sign of the difference) leads to a decrease of forecasting performance. It reduces the informational content in data which is key for signal extraction.

References

Alzaid, A. and M. A. Omair (2010). On the Poisson difference distribution inference and applications. *Bulletin of the Malaysian Mathematical Sciences Society 33(1)*, 17–45.

Blasques, F., S. J. Koopman, and A. Lucas (2015). Information theoretic optimality of observation driven time series models for continuous responses. *Biometrika 102(2)*, 325–343.

Cattelan, M., C. Varin, and D. Firth (2013). Dynamic Bradley-Terry modelling of sports tournaments. *Applied Statistics 62(1)*, 135–150.

Constantinou, A. C. and N. E. Fenton (2012). Solving the problem of inadequate scoring rules for assessing probabilistic football forecast models. *Journal of Quantitative Analysis in Sports 8(1)*, 1.

Creal, D. D., S. J. Koopman, and A. Lucas (2013). Generalized autoregressive score models with applications. *Journal of Applied Econometrics 28(5)*, 777–795.

Crowder, M., M. J. Dixon, A. Ledford, and M. Robinson (2002). Dynamic modelling and prediction of English football league matches for betting. *The Statistician 51(2)*, 157–168.

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.

Dixon, M. J. and M. E. Robinson (1998). A birth process model for association football matches. *The Statistician 47(3)*, 523–538.

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

Fahrmeir, L. and G. Tutz (1994). Dynamic stochastic models for time-dependent ordered paired comparison systems. *J. American Statistical Association 89(428)*, 1438–1449.

Famoye, F. (2010). On the bivariate negative binomial regression model. *Journal of Applied Statistics 37(6)*, 969–981.

Forrest, D. and R. Simmons (2000). Making up the results: The work of the football pools panel, 1963-1997. *The Statistician 49(2)*, 253–260.

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

Goddard, J. and I. Asimakopoulos (2004). Forecasting football results and the efficiency of fixed-odds betting. *Journal of Forecasting 23*, 51–66.

Gourieroux, C., A. Monfort, and A. Trognon (1984). Pseudo maximum likelihood methods: Applications to Poisson models. *Econometrica 52(3)*, 701–720.

Greene, W. H. (2012). *Econometric Analysis*. New York: Pearson Education.

Holgate, P. (1964). Estimation for the bivariate Poisson distribution. *Biometrika 51(1)*, 241–245.

Hvattum, L. M. and H. Arntzen (2010). Using Elo ratings for match result prediction in association football. *International Journal of Forecasting 26*, 460–470.

Irwin, J. O. (1937). The frequency distribution of the difference between two independent variates following the same Poisson distribution. *J. Royal Statistical Society A 100(3)*, 415–416.

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

Karlis, D. and I. Ntzoufras (2003). Analysis of sports data by using bivariate Poisson models. *The Statistician 52(3)*, 381–393.

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

Knorr-Held, L. (2000). Dynamic rating of sports teams. *The Statistician 49(2)*, 261–276.

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

Kocherlakota, S. and K. Kocherlakota (2001). Regression in the bivariate Poisson distribution. *Communications in Statistics - Theory and Methods 30(5)*, 815–825.

Koning, R. H. (2000). Balance in competition in Dutch soccer. *The Statistician 49(3)*, 419–431.

Koopman, S. J. and R. Lit (2014). A dynamic bivariate Poisson model for analysing and forecasting match results in the English Premier League. *J. Royal Statistical Society A 178(1)*, 167–186.

Koopman, S. J., R. Lit, and A. Lucas (2017). Intraday stochastic volatility in discrete price changes: the dynamic Skellam model. *J. American Statistical Association 102*, forthcoming.

Koopman, S. J., A. Lucas, and M. Scharth (2016). Predicting time-varying parameters with parameter-driven and observation-driven models. *Review of Economics and Statistics 98(1)*, 97–110.

Lit, R. (2016). *Time-varying parameter models for discrete valued time series*. Number 642 in Tinbergen Institute Research Series, Ph.D. dissertation. Amsterdam: Thela Thesis and Tinbergen Institute.

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

Rue, H. and O. Salvesen (2000). Prediction and retrospective analysis of soccer matches in a league. *The Statistician 49(3)*, 399–418.

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

Table 2

Rank Probability Scores and Diebold Mariano test results

We present probability forecast losses based on the average rank probability score (ARPS) obtained by the forecasting of seven seasons of match results for each of six European football competitions with an expanding window. The RPS are averaged over the out-of-sample window to get a summarizing statistic. The Diebold-Mariano (DM) statistic is based on the whole out-of-sample window with one-step-ahead forecasts. The lowest ARPS for each column is darkly coloured (in blue). The column title '# par' denotes the dimension of the parameter vector ψ or ψ^* where N is the number of teams in the data set (all seasons combined). The DM statistic is asymptotically distributed as a standard normal random variable and hence rejects the null hypothesis of equal predictive accuracy at the 5% level of significance if the DM test statistic is smaller than −1.96 (the benchmark model performs significantly worse) or larger than 1.96 (the benchmark model performs significantly better). The benchmark model is the best performing model per competition. The column 'Time' denotes average computer time in seconds for maximizing the log-likelihood function (averaged over the six competitions). Computations are performed on a i7-2600, 3.40 GHz desktop PC using four cores.

		Additional			England		Germany		Spain	
Distribution	Type	Specifications	Time	# par	ARPS	DM	ARPS	DM	ARPS	DM
Biv. Poisson	Static	$\sum_{i=1}^{N} \alpha_i = 0$	31.20	$2N+2$	0.2062	6.17	0.2165	4.76	0.1980	5.98
Skellam	Static	$\sum_{i=1}^{N} \alpha_i = 0$	30.95	$2N+1$	0.2068	6.68	0.2167	4.81	0.1983	6.06
Ordered probit	Static	$\sum_{i=1}^{N} \alpha_i = 0$	6.68	$N+2$	0.2083	6.69	0.2173	5.07	0.1989	5.74
Biv. Poisson	Semi-dynamic	$\sum_{i=1}^{N} \alpha_i = 0$	116.47	$2N+2$	0.2014	4.15	0.2127	3.35	0.1942	4.28
Skellam	Semi-dynamic	$\sum_{i=1}^{N} \alpha_i = 0$	65.53	$2N+1$	0.2034	5.55	0.2156	4.47	0.1944	3.88
Ordered probit	Semi-dynamic	$\sum_{i=1}^{N} \alpha_i = 0$	40.56	$N+2$	0.2027	4.39	0.2138	3.88	0.1952	4.01
Biv. Poisson	Dynamic	$f_0 \in \psi^*$	183.30	$2N+6$	0.2024	4.69	0.2126	2.99	0.1952	4.54
Skellam	Dynamic	$f_0 \in \psi^*$	252.45	$2N+5$	0.2020	4.69	0.2120	2.54	0.1953	4.34
Ordered probit	Dynamic	$f_0 \in \psi^*$	41.65	$N+4$	0.2030	3.93	0.2140	3.27	0.1962	4.61
Biv. Poisson	Dynamic	$b_1 = b_2 = 1,\ f_0$ static init	1.72	4	0.1987	2.44	0.2095	2.15	0.1915	-
Skellam	Dynamic	$b_1 = b_2 = 1,\ f_0$ static init	1.30	3	0.1982	1.01	0.2092	0.73	0.1918	1.03
Ordered probit	Dynamic	$b_1 = 1,\ f_0$ static init	0.80	3	0.2001	3.06	0.2104	2.67	0.1929	3.09
Biv. Poisson	Dynamic	f_0 static init	8.44	6	0.1984	1.22	0.2089	-	0.1916	1.57
Skellam	Dynamic	f_0 static init	3.76	5	0.1981	-	0.2089	0.01	0.1918	1.03
Ordered probit	Dynamic	f_0 static init	1.30	4	0.1996	2.32	0.2103	2.51	0.1929	3.09
Biv. Poisson	Dynamic	State Space Model	1 hour	6	0.1984	1.26	0.2099	2.10	0.1928	3.04

Table 3

Rank Probability Scores and Diebold Mariano test results

The table presents losses based on the average rank probability score (ARPS) obtained by the forecasting of seven seasons of match results for each of six European football competitions with an expanding window. The RPS are averaged over the out-of-sample window to get a summarizing statistic. The Diebold Mariano statistic is based on the whole out-of-sample window with one-step-ahead forecasts. The lowest ARPS for each column is colored in blue. The column title '# par' denotes the dimension of the parameter vector ψ^* where N is the number of teams in the data set (all seasons combined). The DM statistic represents the Diebold and Mariano (1995) statistic which is asymptotically distributed as a standard normal random variable and hence rejects the null hypothesis of equal predictive accuracy at the 5% level of significance if the DM test statistic is smaller than −1.96 (the benchmark model performs significantly worse) or larger than 1.96 (the benchmark model performs significantly better). The benchmark model is the best performing model per competition. The column 'Time' denotes the time in seconds it took to maximize the likelihood (averaged over the six competitions). Computations are performed on a i7-2600, 3.40 GHz desktop PC using four cores.

		Additional			Italy		France		Netherlands	
Distribution	Type	Specifications	Time	# par	ARPS	DM	ARPS	DM	ARPS	DM
Biv. Poisson	Static	$\sum_{i=1}^{N} \alpha_i = 0$	31.20	$2N+2$	0.2042	3.44	0.2109	4.34	0.1974	2.34
Skellam	Static	$\sum_{i=1}^{N} \alpha_i = 0$	30.95	$2N+1$	0.2049	3.77	0.2113	4.56	0.1976	2.41
Ordered probit	Static	$\sum_{i=1}^{N} \alpha_i = 0$	6.68	$N+2$	0.2050	3.82	0.2126	5.23	0.1973	2.29
Biv. Poisson	Semi-dynamic	$\sum_{i=1}^{N} \alpha_i = 0$	116.47	$2N+2$	0.2025	2.78	0.2070	1.28	0.1939	0.58
Skellam	Semi-dynamic	$\sum_{i=1}^{N} \alpha_i = 0$	65.53	$2N+1$	0.2036	3.45	0.2077	2.02	0.1944	0.88
Ordered probit	Semi-dynamic	$\sum_{i=1}^{N} \alpha_i = 0$	40.56	$N+2$	0.2033	3.12	0.2086	2.66	0.1943	1.05
Biv. Poisson	Dynamic	$f_0 \in \psi^*$	183.30	$2N+6$	0.2028	2.60	0.2082	2.55	0.1949	1.17
Skellam	Dynamic	$f_0 \in \psi^*$	252.45	$2N+5$	0.2027	2.42	0.2081	2.51	0.1942	0.70
Ordered probit	Dynamic	$f_0 \in \psi^*$	41.65	$N+4$	0.2031	2.69	0.2093	3.41	0.1938	0.43
Biv. Poisson	Dynamic	$b_1 = b_2 = 1,\ f_0$ static init	1.72	4	0.2011	2.27	0.2066	1.76	0.1940	0.73
Skellam	Dynamic	$b_1 = b_2 = 1,\ f_0$ static init	1.30	3	0.2008	0.80	0.2062	1.31	0.1942	1.07
Ordered probit	Dynamic	$b_1 = 1,\ f_0$ static init	0.80	3	0.2023	2.89	0.2071	2.21	0.1938	1.08
Biv. Poisson	Dynamic	f_0 static init	8.44	6	0.2005	-	0.2060	0.37	0.1936	0.31
Skellam	Dynamic	f_0 static init	3.76	5	0.2007	0.66	0.2059		0.1941	1.01
Ordered probit	Dynamic	f_0 static init	1.30	4	0.2019	2.22	0.2067	1.59	0.1934	-
Biv. Poisson	Dynamic	State Space Model	1 hour	6	0.2010	1.84	0.2066	1.62	0.1938	0.45

Appendices

A First derivatives of probability mass functions

For notational simplicity, throughout this appendix the subscripts that denote time t and teams i, j are suppressed for intensities and observations.

A.1 Double Poisson distribution

A pair of counts (X, Y) which is assumed to be distributed by the double Poisson distribution with intensities $\lambda_1, \lambda_2 > 0$ has probability mass function

$$p(x, y; f_t, \psi^*) = \frac{e^{-\lambda_1}\lambda_1^x}{x!} \times \frac{e^{-\lambda_2}\lambda_2^y}{y!}. \tag{A.1}$$

The mean and variance of the marginals are $\mathbb{E}(X) = \mathbb{V}\mathrm{ar}(X) = \lambda_1$ and $\mathbb{E}(Y) = \mathbb{V}\mathrm{ar}(Y) = \lambda_2$. If the intensities are functions of latent strengths of attack and defence as in (3), the first derivative of the double Poisson distribution with respect to $f_{ijt} = (\alpha_{it}, \alpha_{jt}, \beta_{it}, \beta_{jt})'$ is given by the 4×1 score vector

$$\nabla_{ijt} = \frac{\partial \log p(x, y|f_t, \mathcal{F}_t, \psi^*)}{\partial f_{ijt}} = (x - \lambda_1, y - \lambda_2, \lambda_2 - y, \lambda_1 - x)'. \tag{A.2}$$

A.2 Bivariate Poisson distribution

The first derivative of the bivariate Poisson mass function in (1) with respect to f_{ijt} is

$$\nabla_{ijt} = \frac{\partial \log p(y_{ijt}|f_t, \mathcal{F}_t, \psi^*)}{\partial f_{ijt}} = \begin{pmatrix} x-\lambda_1-U(f_t,\psi^*) \\ y-\lambda_2-U(f_t,\psi^*) \\ \lambda_2-y+U(f_t,\psi^*) \\ \lambda_1-x+U(f_t,\psi^*) \end{pmatrix}, \tag{A.3}$$

where $U(f_t, \psi^*) = S(1, f_t, \psi^*)/S(0, f_t, \psi^*)$ with

$$S(q, f_t, \psi^*) = \sum_{k=0}^{min(x,y)} \binom{x}{k}\binom{y}{k} k!\, k^q \left(\frac{\lambda_3}{\lambda_1\,\lambda_2}\right)^k, \qquad q = 0, 1.$$

We notice that $S(1, f_t, \psi^*) = 0$ when $\lambda_3 = 0$ and $S(0, f_t, \psi^*) = 1$ when $min(x, y) = 0$ so that function $U(f_t, \psi^*)$ is properly defined for all $\lambda_3 \geq 0$, see also the online appendix of Koopman and Lit (2014). Finally we observe that for $\lambda_3 = 0$ the score vector of the bivariate Poisson distribution reduces to the score vector of the double Poisson distribution in equation (A.2).

A.3 Skellam distribution

The first derivative of the Skellam mass function in (5) with respect to $f_{ijt} = (\alpha_{it}, \alpha_{jt}, \beta_{it}, \beta_{jt})'$ is given by the 4×1 score vector

$$\nabla_{ijt} = \frac{\partial \log\, p(y_{ijt}|f_t, \mathcal{F}_t, \psi^*)}{\partial f_{ijt}} = \begin{pmatrix} z-\lambda_1 & +W(f_t,\psi^*) \\ -\lambda_2 & +W(f_t,\psi^*) \\ \lambda_2 & -W(f_t,\psi^*) \\ -z+\lambda_1 & -W(f_t,\psi^*) \end{pmatrix}, \tag{A.4}$$

where $W(f_t, \psi^*)$ is defined as

$$\sqrt{\lambda_1\lambda_2} \times \frac{I_{z+1}(2\sqrt{\lambda_{1,ijt}\lambda_2})}{I_z(2\sqrt{\lambda_1\lambda_2})}.$$

A.4 Ordered probit distribution

The first derivative of $p(C; f_t, \psi^*)$ with respect to $f_{ijt} = (\gamma_{it}, \gamma_{jt})'$ is given by the 2×1 score vector

$$\nabla_{ijt} = \frac{\partial \log\, p(y_{ijt}|f_t, \mathcal{F}_t; \psi^*)}{\partial f_{ijt}} = \begin{cases} \begin{pmatrix} \frac{\phi(\kappa_2-\lambda_{6,ijt})}{1-\Phi(\kappa_2-\lambda_{6,ijt})} \\ \frac{-\phi(\kappa_2-\lambda_{6,ijt})}{1-\Phi(\kappa_2-\lambda_{6,ijt})} \end{pmatrix} & \text{if } C = 2, \\ \begin{pmatrix} \frac{\phi(\kappa_1-\lambda_{6,ijt})-\phi(\kappa_2-\lambda_{6,ijt})}{\Phi(\kappa_2-\lambda_{6,ijt})-\Phi(\kappa_1-\lambda_{6,ijt})} \\ \frac{\phi(\kappa_2-\lambda_{6,ijt})-\phi(\kappa_1-\lambda_{6,ijt})}{\Phi(\kappa_2-\lambda_{6,ijt})-\Phi(\kappa_1-\lambda_{6,ijt})} \end{pmatrix} & \text{if } C = 1, \\ \begin{pmatrix} \frac{-\phi(\kappa_1-\lambda_{6,ijt})}{\Phi(\kappa_1-\lambda_{6,ijt})} \\ \frac{\phi(\kappa_1-\lambda_{6,ijt})}{\Phi(\kappa_1-\lambda_{6,ijt})} \end{pmatrix} & \text{if } C = 0, \end{cases} \tag{A.5}$$

where $\phi(\cdot)$ is the standard normal pdf.

B More Tables and Figures

Table B.1

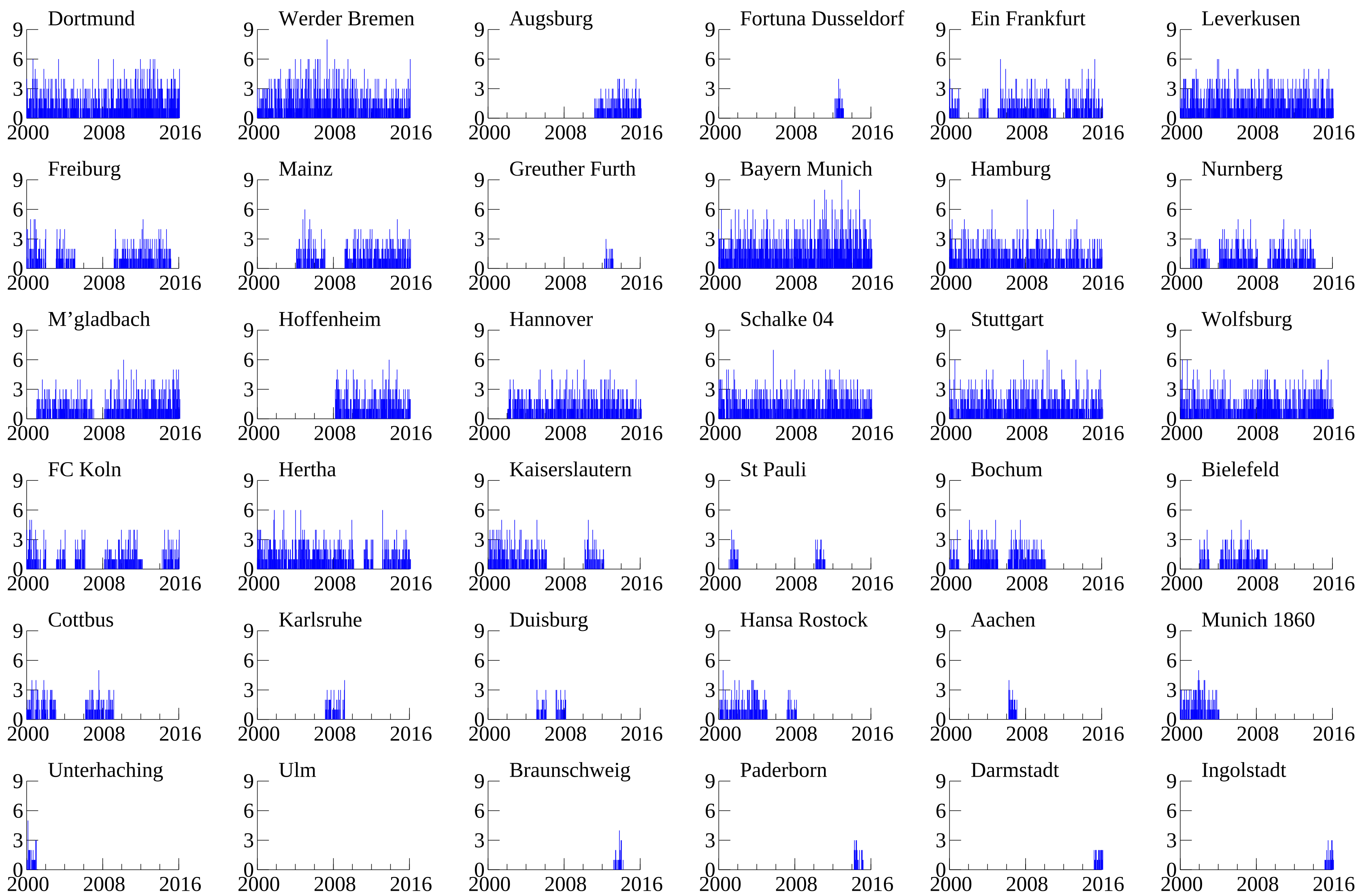
Main contributions to the football literature

The main contributions in the literature on modelling football match results are organized into nine categories. The three columns are for the dynamic extension of the model: 'Static' means no dynamics at all, 'Dynamic' means fully dynamic and 'Semi-dynamic' refers to the method of weighted maximum likelihood estimation. The three rows are for the type of observation in which the match result is measured: a pairwise observation ('Goals'), a difference between the number of goals ('Difference'), or a category variable ('Toto'). Although it is not the main focus or contribution of this paper, we address the empty square in the middle of this table by considering weighted maximum likelihood estimation for the parameters in the static Skellam model. Our main contribution is for the last column as we propose an alternative dynamic extension to the three observation densities.

	Static	Semi-dynamic	Dynamic
Goals	Maher (1982) Karlis and Ntzoufras (2003) Goddard (2005) Dixon and Robinson (1998)	Dixon and Coles (1997)	Crowder et al. (2002) Rue and Salvesen (2000) Koopman and Lit (2014)
Difference	Karlis and Ntzoufras (2009)		Lit (2016, Ch. 4)
Toto	Goddard et al. (2004) Forrest and Simmons (2000) Koning (2000)	Cattelan et al. (2013)	Fahrmeir and Tutz (1994) Knorr-Held (2000) Hvattum and Arntzen (2010)

Figure B.1

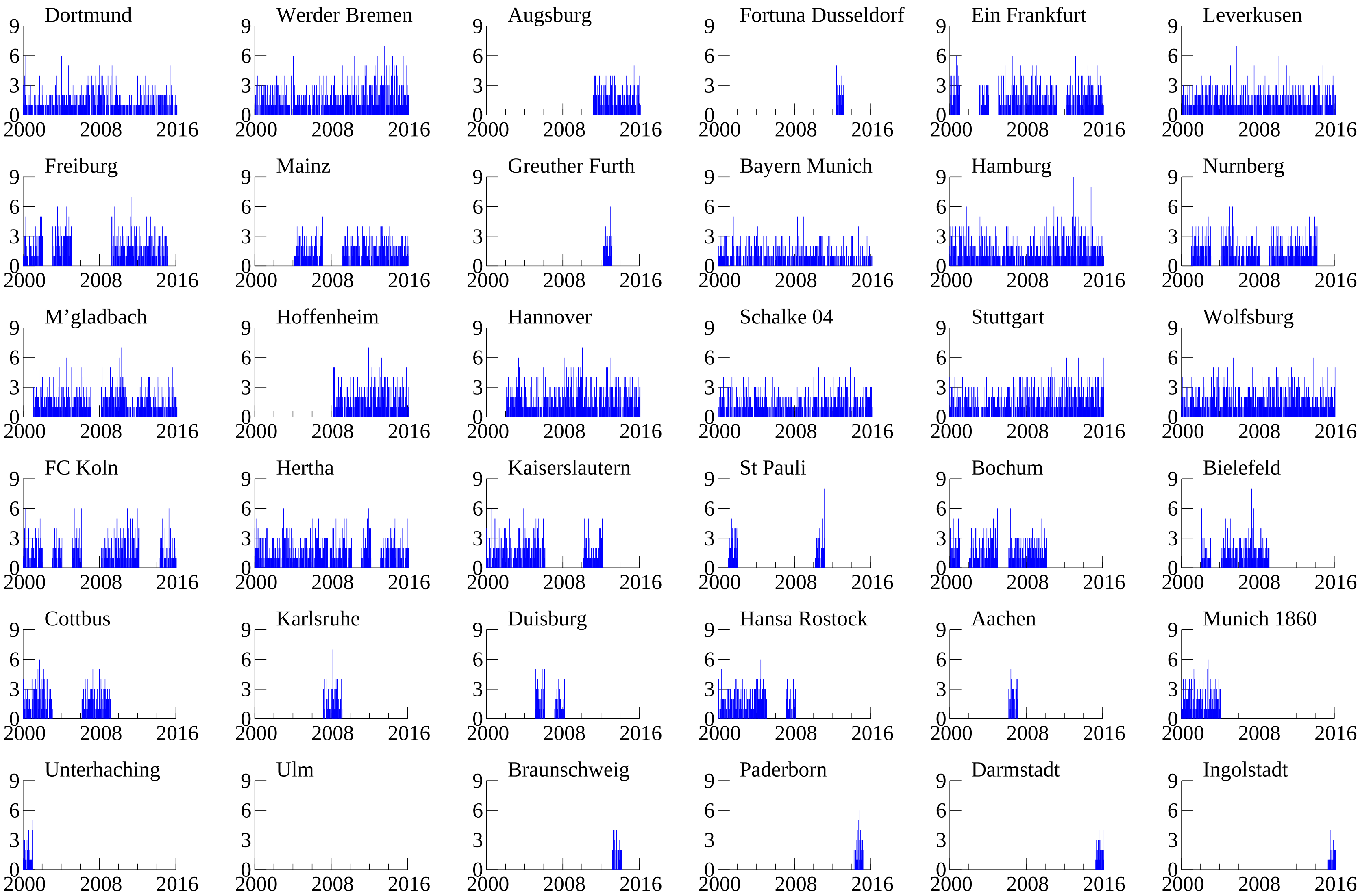
Scored goals in the German Bundesliga between 2000 and 2016



The figure presents the number of goals scored in the German Bundesliga for the period 2000-2016. The gaps in the data are due to promotion and relegation and several teams only played for one season after which they immediately relegated back to a lower division. The panel of Ulm is empty. This is because Ulm played in the 1999-2000 season, which was the initializing season for the dynamic model, but did not play in the other seasons.

Figure B.2

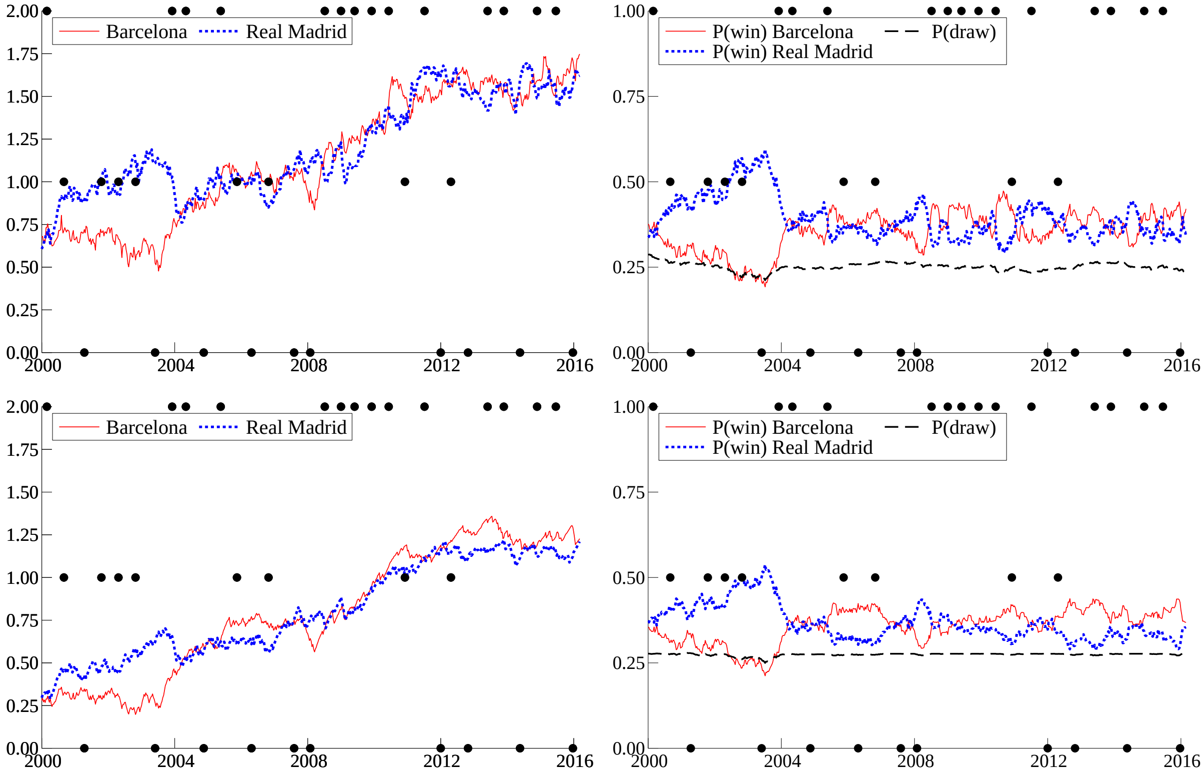
Conceded goals in the German Bundesliga between 2000 and 2016



The figure presents the number of goals conceded in the German Bundesliga for the period 2000-2016. The gaps in the data are due to promotion and relegation and several teams only played for one season after which they immediately relegated back to a lower division. The panel of Ulm is empty. This is because Ulm played in the 1999-2000 season, which was the initializing season for the dynamic model, but did not play in the other seasons.

Figure B.3

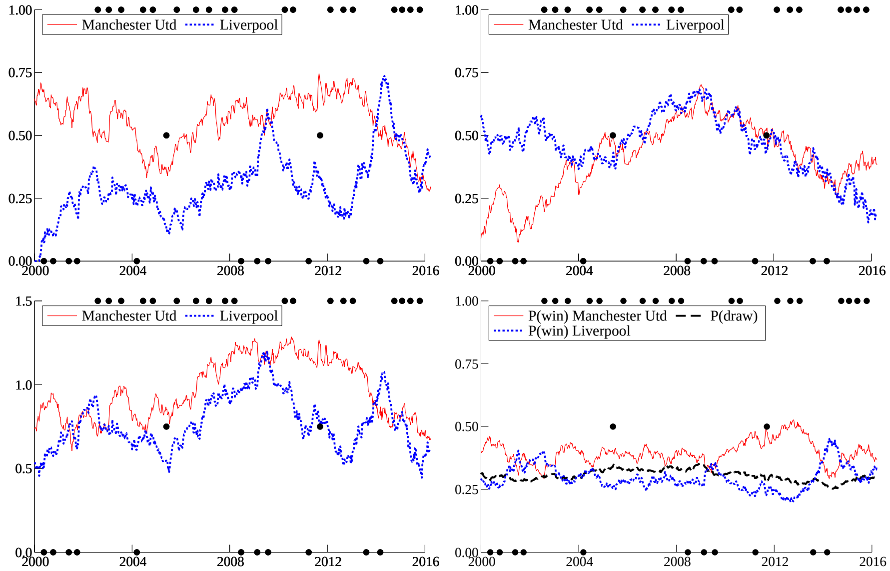
Extracted strengths of Barcelona and Real Madrid :
Skellam and ordered probit models



All panels: a dot on top of the panel represents a win for Barcelona, dot at the bottom is a win for Real Madrid, and a draw is represented by a dot in the middle of the panel. Top left panel: sum of extracted strengths of attack and defence from dynamic Skellam model. Top right panel: probability of toto results from the dynamic Skellam model. Bottom left panel: sum of extracted strengths of attack and defence from the dynamic ordered probit model. Bottom right panel: probability of toto results from the dynamic ordered probit model.

Figure B.4

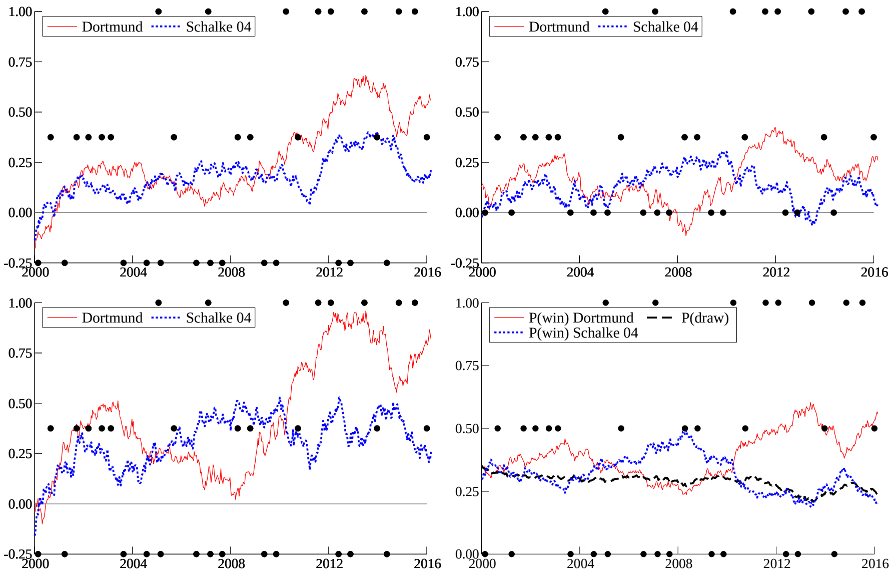
Extracted strengths of Manchester United and Liverpool



All panels: a dot on top of the panel represents a win for Manchester United, dot at the bottom is a win for Liverpool, and a draw is represented by a dot in the middle of the panel. Top left panel: time series plot of extracted strengths of attack. Top right panel: time series plot of extracted strengths of defence. Bottom left panel: sum of extracted strengths of attack and defence. Bottom right panel: probability of toto results from the dynamic Bivariate Poisson model.

Figure B.5

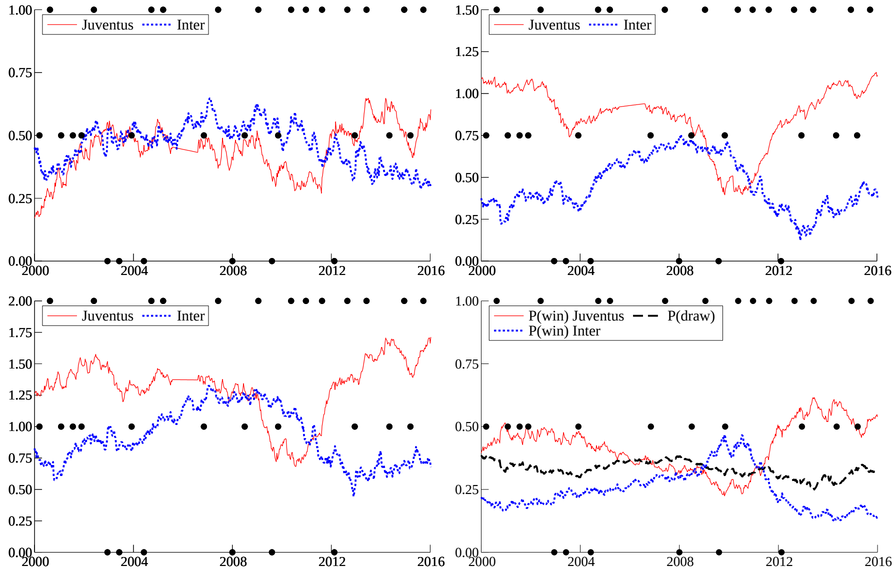
Extracted strengths of Dortmund and Schalke 04



All panels: a dot on top of the panel represents a win for Dortmund, dot at the bottom is a win for Schalke 04, and a draw is represented by a dot in the middle of the panel. Top left panel: time series plot of extracted strengths of attack. Top right panel: time series plot of extracted strengths of defence. Bottom left panel: sum of extracted strengths of attack and defence. Bottom right panel: probability of toto results from the dynamic Bivariate Poisson model.

Figure B.6

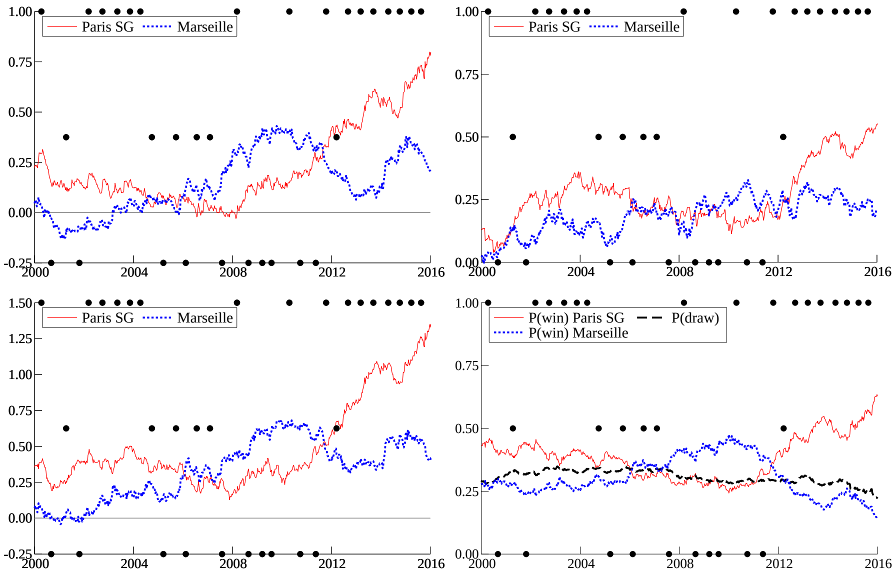
Extracted strengths of Juventus and Inter Milan



All panels: a dot on top of the panel represents a win for Juventus, dot at the bottom is a win for Inter Milan, and a draw is represented by a dot in the middle of the panel. Top left panel: time series plot of extracted strengths of attack. Top right panel: time series plot of extracted strengths of defence. Bottom left panel: sum of extracted strengths of attack and defence. Bottom right panel: probability of toto results from the dynamic Bivariate Poisson model.

Figure B.7

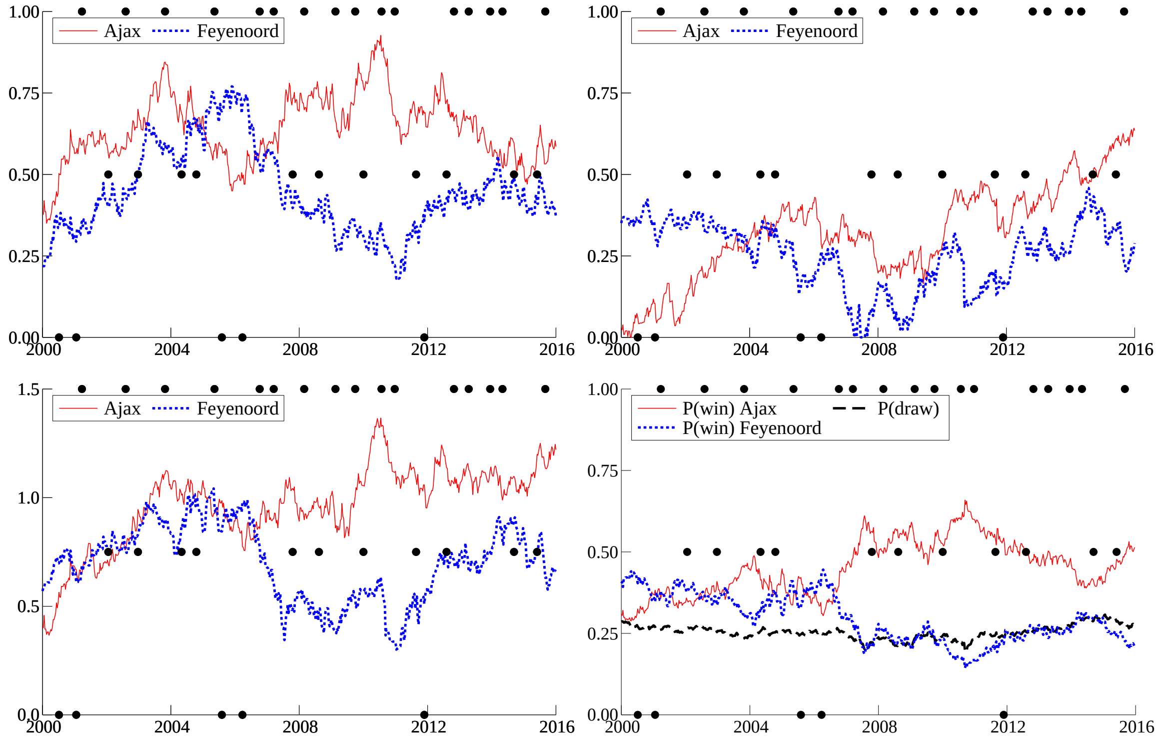
Extracted strengths of Paris SG and Marseille



All panels: a dot on top of the panel represents a win for Paris SG, dot at the bottom is a win for Marseille, and a draw is represented by a dot in the middle of the panel. Top left panel: time series plot of extracted strengths of attack. Top right panel: time series plot of extracted strengths of defence. Bottom left panel: sum of extracted strengths of attack and defence. Bottom right panel: probability of toto results from the dynamic Bivariate Poisson model.

Figure B.8

Extracted strengths of Ajax and Feyenoord



All panels: a dot on top of the panel represents a win for Ajax, dot at the bottom is a win for Feyenoord, and a draw is represented by a dot in the middle of the panel. Top left panel: time series plot of extracted strengths of attack. Top right panel: time series plot of extracted strengths of defence. Bottom left panel: sum of extracted strengths of attack and defence. Bottom right panel: probability of toto results from the dynamic Bivariate Poisson model.