

## Dynamic Bayesian forecasting models of football match outcomes with estimation of the evolution variance parameter

ALUN OWEN

*Mathematics Education Centre, Loughborough University, Loughborough,  
Leicestershire LE11 3TU, UK*

a.j.owen@lboro.ac.uk

[Received on 29 September 2009; accepted on 29 November 2010]

Statistical models of football (soccer) match outcomes have potential applications to areas such as the development of team rankings and football betting markets. Much of the published work in this context has typically focused on the use of generalized linear models, which are non-dynamic in the sense that the parameters in the model, which often represent the underlying abilities of each team, are assumed to remain constant over time. Dynamic generalized linear models (DGLMs) on the other hand allow the abilities of each team to vary over time. This paper illustrates the application of a DGLM in the context of football match outcome prediction and describes improvements on similar work previously presented by the author, in relation to the estimation of a parameter in the model, referred to as the evolution variance, which is crucial in terms of optimizing the predictive performance of these types of models. Match results data from the Scottish Premier League from 2003/2004 to 2005/2006 are used to show that the DGLM approach provides improved predictive probabilities of future match outcomes compared to the non-dynamic form of the model. DGLMs are also Bayesian in terms of their structure and so a Bayesian approach to parameter estimation is required. This paper therefore illustrates a practical implementation of the DGLM model that can easily be deployed using the freely available software WinBUGS.

*Keywords:* dynamic generalized linear models; Bayesian; evolution variance football; Scottish Premier League.

### 1. Introduction

Much of the interest in the development of statistical models of football (soccer) data typically relates to the development of either team rankings or more commonly predictive probabilities of future match outcomes, which are typically considered in terms of a home win, draw or away win. This latter application is perhaps more popular because of their obvious applications to football betting markets. The published literature with regard to modelling probabilities of future match outcomes has often considered models from a classical standpoint, making use of the generalized linear modelling (GLM) framework that uses maximum likelihood methods for parameter estimation. One problem with this approach is that the parameters in the model are typically assumed to remain constant over time, which would seem unrealistic given the potential for variability in a team's performance over time.

Owen (2009) presented an alternative approach to the development of such models in the context of football match outcomes, which allows some or all of the parameters in the model to be time dependent. This was based on the use of the dynamic generalized linear modelling (DGLM) framework described in West & Harrison (1997). However, Owen (2009) also illustrated that one of the most important considerations with the DGLM approach in this context, related to the need to model the temporal changes in team performances, via suitable estimation of a parameter in the model referred to as the

evolution variance. Owen (2009) identified problems with estimating the evolution variance if this was retained as a parameter in the model and instead estimated this by optimizing the predictive performance of the model. This paper not only presents a more detailed discussion of the work presented in Owen (2009) but also extends this by illustrating how the results of the work presented there can be used to actually retain the evolution variance as a parameter in the model.

DGLMs are well suited to forecasting applications and so provide a natural approach to model development in our context since we are primarily interested in future match prediction. A DGLM model, where goals scored are assumed to be generated by a Poisson process, is applied to match results data from the Scottish Premier League (SPL) over a period of three seasons from 2003/2004 to 2005/2006. DGLMs are also very much Bayesian in flavour since the parameter estimates derived at any particular point in time effectively serve as prior information for parameter estimation at the next time point. Therefore, the use of a Bayesian approach to model development is typically required to facilitate parameter estimation. Here, we make use of Markov chain Monte Carlo (MCMC) simulation methods to derive parameter estimates and predictive probabilities of future match outcomes, using the freely available software WinBUGS (see <http://www.mrc-bsu.cam.ac.uk/bugs>).

The paper is organized such that Section 2 describes the structure of the model, including consideration of an interesting problem of dynamically modelling parameters that are subject to parameter constraints. An assessment of the model's predictive performance is presented in Section 3, which also includes estimation of the appropriate level of underlying temporal variability in teams abilities. This last issue being a key consideration in the development of these types of dynamic models. The dynamic model presented in this paper has some similarities with an approach described in Rue & Salvesen (2000), and so Section 4 of this paper highlights the differences between the two approaches and illustrates the advantages to be gained using the model presented here. Similar dynamic models are also discussed in Crowder *et al.* (2002), but those authors derive parameter estimates via an approximation method, whereas the aim here was to present an implementation of a dynamic model that can be deployed easily using readily available software. The paper by Dixon & Coles (1997) also attempts to take account of the temporal nature of the data by employing a weighted maximum likelihood approach to parameter estimation, but this is a quite different approach to that being considered here. The work presented in this paper is concluded with some comments and a discussion of proposed future work in Section 5.

## 2. A dynamic generalized linear model

### 2.1 Model specification

The model developed here, which was previously described in Owen (2009), is based on a non-dynamic model originally presented in Maher (1982). This model has also been extensively considered by other authors; see, e.g. Dixon & Coles (1997), Lee (1998) and Karlis & Ntzoufras (2003). However, the model is extended here to the dynamic case, so that the number of goals scored by team  $i$  playing at home and team  $j$  playing away, in a match played at time  $t$ , are denoted by  $X_{i,j,t}$  and  $Y_{i,j,t}$ , respectively, and are modelled as independent Poisson variables as follows:

$$X_{i,j,t} \sim \text{Po}(\mu_{i,j,t}), \quad (1)$$

$$Y_{i,j,t} \sim \text{Po}(\lambda_{i,j,t}), \quad (2)$$

$$\log(\mu_{i,j,t}) = \alpha_{i,t} + \beta_{j,t} + \gamma_H, \quad (3)$$

$$\log(\lambda_{i,j,t}) = \alpha_{j,t} + \beta_{i,t} + \gamma_A, \quad i, j = 1, 2, \dots, n, \quad t = 1, 2, \dots, T, \quad (4)$$

with  $n$  teams playing  $T$  rounds of matches. The  $\alpha_{i,t}$  and  $\beta_{i,t}$  measure the attack and defence abilities, respectively, of team  $i$  at time  $t$  and are the same irrespective of whether a team is playing at home or away.

To ensure unique identifiability of the parameters, two constraints are required and specified here as

$$\sum_{i=1}^n \alpha_{i,t} = 0, \quad (5)$$

$$\sum_{i=1}^n \beta_{i,t} = 0. \quad (6)$$

The attack and defence parameters,  $\alpha_{i,t}$  and  $\beta_{i,t}$ , therefore represent the attacking and defensive strengths, relative to an average team that has average attack and defence parameters of 0. Because of this, the parameters  $\gamma_H$  and  $\gamma_A$  reflect the underlying (natural logarithm of) overall average scoring rates at home and away, respectively, and are assumed to remain constant over time.

The parameterization given by (3) and (4) differs slightly from that used by some previous authors, where a home effect parameter  $\gamma$  is included in (3), but  $\gamma_A$  does not appear in (4), and required just a single identifiability constraint. However, the parameterization given by (3) and (4) has an attractive symmetrical property since the attack and defensive parameters are treated equally. This proves to be very useful when we consider the dynamic aspects of the model. In addition, when using MCMC simulation methods, the convergence properties of the simulated values for the defence parameters are improved through the use of our parameterization, which is illustrated later in Section 2.3. Note that the difference  $\gamma_H - \gamma_A$  is in fact equivalent to the home effect,  $\gamma$ , used by others, and so a home effect is indeed present in our parameterization and can easily be estimated if required.

The terms specified in (1)–(4) form the ‘observation’ component of the model, which provide the basis of the usual GLM. To fully specify this as a dynamic model, an ‘evolution’ component is also required, which describes the stochastic behaviour of the time-dependent parameters. Here, the evolution component is specified as a random walk for both the attack and defence parameters as follows:

$$\alpha_{i,t} \sim N(\alpha_{i,t-1}, \sigma^2), \quad (7)$$

$$\beta_{i,t} \sim N(\beta_{i,t-1}, \sigma^2). \quad (8)$$

In (7) and (8), the parameter,  $\sigma^2$ , represents the ‘evolution variance’, which for simplicity, is assumed to remain constant over time and to be common to all teams. In addition, the evolution variance is assumed to be common to both the attack and defence parameters. This later assumption is of course only made possible through the use of the parameterization in (3) and (4) since this treats the attack and defence parameters equally. This use of a common evolution variance is advantageous in a number of ways, most notably that it keeps the number of parameters in an already highly parameterized model to a minimum.

To complete the specification of the model, priors are specified as follows:

$$\alpha_{i,0} \sim N(m_{\alpha_i}, \sigma_0^2), \quad (9)$$

$$\beta_{i,0} \sim N(m_{\beta_i}, \sigma_0^2), \quad (10)$$

$$\gamma'_H \sim \text{Ga}(g_H, h_H), \quad \gamma'_A \sim \text{Ga}(g_A, h_H), \quad (11)$$

where  $\gamma_H = \log(\gamma'_H)$ ,  $\gamma_A = \log(\gamma'_A)$ , the  $\alpha_{i,0}$  and  $\beta_{i,0}$  represent baseline attack and defence strengths at the beginning of a season prior to any matches being played and the  $m_\alpha$  and  $m_\beta$  are known constants representing prior means for these baseline attack and defence strengths. The parameter  $\sigma_0^2$  represents the ‘prior variance’, which, again for simplicity, is also assumed to be common to all teams and to both the attack and defence parameters.

## 2.2 MCMC sampling with identifiability constraints using WinBUGS

Parameter estimation was facilitated via a MCMC sampling approach using WinBUGS. However, sampling the  $\alpha_{i,t}$  and  $\beta_{i,t}$  directly is problematic since these need to be sampled according to the evolution relationship given by (7) and (8), but in a manner such that the identifiability constraints (5) and (6) hold for all  $t$ . Ideas taken from West & Harrison (1997) and Knorr-Held (2000) are used to overcome this problem. Firstly, to make the notation easier, for the attack parameters, we define

$$\mathbf{a}_t = [\alpha_{1,t}, \alpha_{2,t}, \dots, \alpha_{n,t}]^T, \quad (12)$$

$$\mathbf{m}_\alpha = [m_{\alpha_1}, m_{\alpha_2}, \dots, m_{\alpha_n}]^T. \quad (13)$$

The evolution of the attack parameters (7) can then be expressed as  $\mathbf{a}_t \sim N(\mathbf{a}_{t-1}, \mathbf{W})$ , where  $\mathbf{W}$  is a diagonal evolution variance matrix with entries  $\sigma^2$ . Similarly, the initial priors on the attack parameters (9) can be expressed as  $\mathbf{a}_0 \sim N(\mathbf{m}_\alpha, \mathbf{W}_0)$ , where  $\mathbf{W}_0$  is a diagonal prior variance matrix with entries  $\sigma_0^2$ . It can then be shown that the identifiability constraint (5) will hold for all  $t$  if  $\mathbf{1}_n^T \mathbf{m}_\alpha = 0$ , where  $\mathbf{1}_n$  is the  $n \times 1$  matrix such that  $\mathbf{1}_n^T = [1, 1, \dots, 1]$  and if the evolution variance matrices,  $\mathbf{W}$  and  $\mathbf{W}_0$ , are modified to variance–covariance matrices  $\mathbf{R}$  and  $\mathbf{R}_0$ , respectively, as follows:

$$\mathbf{R} = \frac{n\sigma^2}{(n-1)} \left( \mathbf{I}_n - \frac{1}{n} \mathbf{1}_n \mathbf{1}_n^T \right), \quad (14)$$

$$\mathbf{R}_0 = \frac{n\sigma_0^2}{(n-1)} \left( \mathbf{I}_n - \frac{1}{n} \mathbf{1}_n \mathbf{1}_n^T \right), \quad (15)$$

where  $\mathbf{I}_n$  is the  $n \times n$  identity matrix. Note that the multiplying factor  $n/(n-1)$  is incorporated so that the variances on the diagonals reflect the variances that are originally specified.

However, the above changes to the variance–covariance structure, given by (14) and (15), present a new problem since  $\mathbf{R}$  and  $\mathbf{R}_0$  are not of full rank and hence have no inverse. As a result, WinBUGS cannot be used directly to sample from a multivariate normal with this variance–covariance structure. This problem could be overcome by sampling from suitable univariate conditional distributions, but this may result in a loss of efficiency, and, given the high number of parameters in the model, it was important to maintain the efficiencies of multivariate sampling as much as possible. A more efficient approach was therefore to sample values for a  $(n-1) \times 1$  vector of unconstrained parameters  $\mathbf{c}_t = [\theta_{1,t}, \theta_{2,t}, \dots, \theta_{n-1,t}]^T$ , from a multivariate normal distribution with zero mean and variance–covariance matrix  $\mathbf{S}_t$  given by

$$\mathbf{S}_t = \frac{n\sigma^2}{(n-1)} (\mathbf{I}_{n-1} + \mathbf{1}_{n-1} \mathbf{1}_{n-1}^T). \quad (16)$$

If a new vector of parameters  $\mathbf{u}_t$  is calculated as  $\mathbf{u}_t = \mathbf{J} \mathbf{c}_t^*$ , where  $\mathbf{c}_t^* = [\theta_{1,t}, \theta_{2,t}, \dots, \theta_{n-1,t}, 0]^T$  and

$$\mathbf{J} = \begin{pmatrix} \mathbf{I}_{n-1} - \frac{1}{n} \mathbf{1}_{n-1} \mathbf{1}_{n-1}^T & \frac{1}{n} \mathbf{1}_{n-1} \\ -\frac{1}{n} \mathbf{1}_{n-1}^T & \frac{1}{n} \end{pmatrix}, \quad (17)$$

it can be shown that the resulting values of  $\alpha_t = \alpha_{t-1} + \mathbf{u}_t$  represent the required sampled values of the attack parameters with the required evolution structure and variance–covariance structure given by (14), and with the identifiability constraint (5) holding for all  $t$ . A similar approach can be applied to the baseline attack parameters, by sampling unconstrained parameters  $\mathbf{c}_0 = [\theta_{1,0}, \theta_{2,0}, \dots, \theta_{n-1,0}]^T$ , from a normal distribution with zero mean and variance–covariance matrix  $\mathbf{S}_0$  given by

$$\mathbf{S}_0 = \frac{n\sigma_0^2}{(n-1)}(\mathbf{I}_{n-1} + \mathbf{1}_{n-1}\mathbf{1}_{n-1}^T) \quad (18)$$

with  $\mathbf{u}_0$  calculated as  $\mathbf{u}_0 = \mathbf{J}\mathbf{c}_0^*$ , where  $\mathbf{c}_0^* = [\theta_{1,0}, \theta_{2,0}, \dots, \theta_{n-1,0}, 0]^T$ , so that  $\alpha_0 = \mathbf{m}_\alpha + \mathbf{u}_0$ .

A similar approach to that described above was also applied to the defence parameters but is not described here for conciseness.

### 2.3 Data and model implementation

Match results data from the SPL over the three seasons from 2003/2004 to 2005/2006 were sourced electronically from [www.football-data.co.uk](http://www.football-data.co.uk). During this period, the SPL had 12 teams each playing 38 rounds of matches per season. The model was retrospectively deployed on a round by round basis, so that retrospective predictive probabilities for the outcomes for each round of matches, in terms of a home win, draw or away win, could be derived, based only on the match results accumulated prior to that round. Each season was analysed separately, with the data for the 2003/2004 season being primarily used for optimization of the evolution variance parameter, and the remaining two seasons to assess the performance of this optimized model.

Prior distributions were derived for all parameters in the model for use at the start of each season, with values for the known constants  $m_\alpha$ ,  $m_\beta$ ,  $g_H$ ,  $h_H$ ,  $g_A$  and  $h_A$ , specified in (9) to (11), being estimated by fitting the non-dynamic form of the model to the previous season's full set of match results. The resulting values for these constants used in the priors for the analysis of each of the three seasons considered are shown in the Appendix. In the SPL, at the end of each season, one team is relegated and replaced by a team promoted from the Scottish League Division 1. As a simplistic approach to the estimation of suitable priors for the promoted team, the priors derived for the attack and defence parameters for the relegated team were utilized here for the promoted team. It is noted that other approaches for the choice of priors and determination of the above known constants are possible, and so further investigation into optimizing this aspect of the modelling process is ongoing. However, no relegation from or promotion to the SPL took place at the start of the 2003/2004 season since the team that should have been promoted (Falkirk) did not satisfy the required criteria with respect to stadium capacity. This was useful in the sense that better prior information was available when analysing the 2003/2004 season since the same 12 teams played in the SPL during the 2002/2003 and the 2003/2004 seasons. Hence, the problem of dealing with choice of prior for the promoted team was not present when analysing the 2003/2004 season, effectively removing this factor from the modelling process.

The sampled values, derived from deploying the model using WinBUGS, typically displayed very good mixing behaviour, and running the sampler for 10,000 iterations, with the first 2,500 iterations being discarded as a burn-in, was assessed as being more than adequate for estimation purposes. The sampler was therefore subsequently run using 10,000 iterations with the last 7,500 iterations being used for estimation purposes. An example of the sample trace (iterations 2,501–5,000) for the latest defence parameter for Celtic, after three rounds of matches had been played during the 2003/2004 season, is shown in Fig. 1(a). This is typical of the sample traces obtained for all parameters in the model. As mentioned previously, the convergence behaviour of the sampled values for the defence parameters was

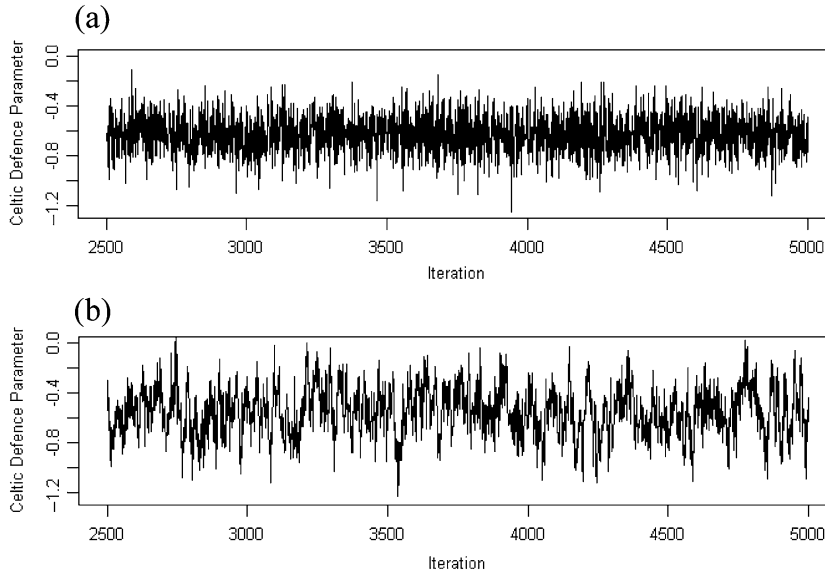


FIG. 1. Trace plots for Celtic defence parameter at round 3 in 2003/2004. Using our parameterization (a) and the alternative parameterization with just a single identifiability constraint (b).

improved through the use of the parameterization adopted here, which treats the attack and defence parameters equally and uses the additional parameter constraint. This improvement is illustrated in Fig. 1(b), which shows the sample trace for the latest Celtic defence parameter, again after three rounds during 2003/2004, but using the alternative parameterization of the model used by others where the additional constraint is not included. In this case, some degree of snaking is evident which indicates that the sampler is not moving quite so freely around the sample space for that parameter.

The evolution variance,  $\sigma^2$ , was initially included as a parameter in the model to be estimated. However, when first considering this parameter, little was known about the likely values and so fairly vague priors were required in this case. Unfortunately, the resulting posterior estimates for  $\sigma^2$  were very sensitive to the choice of prior. Therefore, in order to acquire information about the likely optimum value of  $\sigma^2$ , this was initially pre-specified as a known constant rather than a parameter in the model. The choice of value for  $\sigma^2$  is critical to effectively model the stochastic changes in the attack and defence parameters, and so this was determined primarily by optimizing the model's short-term predictive performance.

Predictive probabilities of future match outcomes, in terms of a home win, draw or away win, were derived retrospectively for matches to be played one-round ahead. This was done by sampling from the predictive distributions of the goals scored for the one-round ahead matches. The proportion of samples where the home goals exceeded the away goals (home win), the home goals equalled the away goals (draw) or the the away goals exceeded the home goals (away win) could then be counted.

One commonly used measure of short-term predictive performance in these types of models, is defined as  $P1$ , which is based on the  $N$  matches played over one complete season as follows:

$$P1 = \exp \left\{ \frac{1}{N} \sum_{k=1}^N \log e[P(O_k)] \right\}, \quad (19)$$

where  $P(O_k)$  represents the one-match ahead predictive probability that match  $k$  would result in the eventual observed outcome,  $O_k$ , of either a 'home win', 'draw' or 'away win'. This is equivalent to the geometric mean of the one-match ahead predictive probabilities for the match outcomes that were actually observed, such that larger values of  $P1$  equate with better predictive performance.

Another short-term predictive performance measure that has been used, e.g. by Knorr-Held (2000), is that defined as  $P2$  as follows:

$$P2 = \frac{1}{N} \sum_{k=1}^N \left\{ [1 - P(O_k)]^2 + P(NO_{1k})^2 + P(NO_{2k})^2 \right\}, \quad (20)$$

where  $P(NO_{1k})$  and  $P(NO_{2k})$  represent the one-match ahead predictive probabilities for the two outcomes (home win, draw or away win) that were not observed in match  $k$ . This is a form of quadratic loss or scoring function and is in effect a discordancy measure or measure of error, such that smaller values of  $P2$  equate with better predictive performance.

Optimization of  $\sigma^2$  was therefore undertaken primarily in relation to maximizing  $P1$  and minimizing  $P2$ . The effect of different values of  $\sigma^2$  on the overall model fit was also investigated through minimization of the deviance information criterion (DIC). The DIC is produced as a standard output by WinBUGS and is considered to be an effective measure of model fit where short-term predictive performance is of interest. In addition, the effect of the choice of the prior variance parameter,  $\sigma_0^2$ , on the predictive performance and model fit measures was also investigated.

### 3. Results

#### 3.1 Estimation of the evolution variance

The effect of the choice of evolution variance,  $\sigma^2$ , on the short-term predictive performance of the model is illustrated in Fig. 2 below. This shows the resulting values of  $P1$  and  $P2$  for various values of  $\sigma^2$ , based on the full set of one-round ahead predictive probabilities for the 2003/2004 season in the SPL.

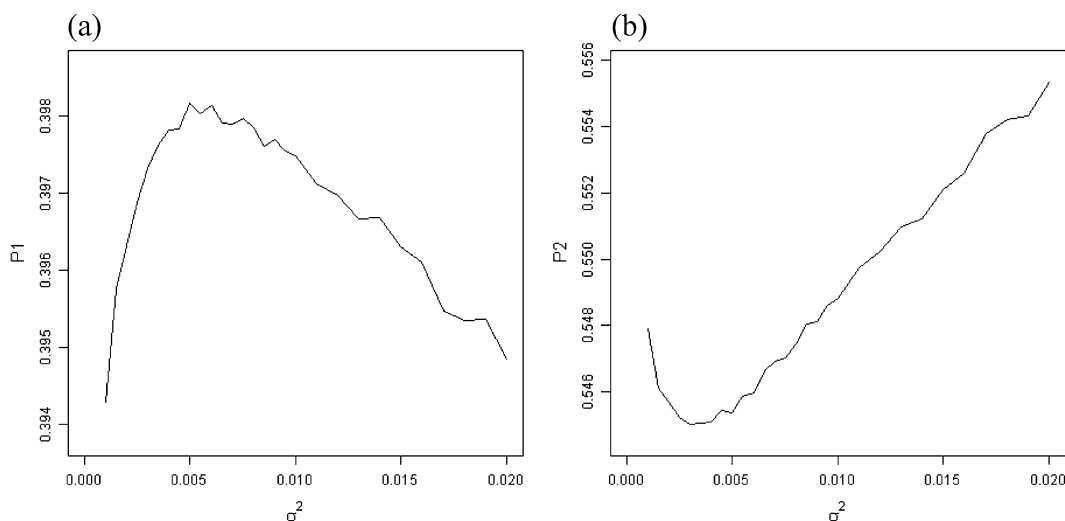


FIG. 2.  $P1$  (a) and  $P2$  (b) versus evolution precision ( $\sigma^2$ ) for 2003/2004.



Figure 2 suggests that  $P1$  is maximized for values of  $\sigma^2$  near 0.004, whereas  $P2$  is minimized for slightly smaller values of  $\sigma^2$  near 0.003.

Assessments based on the DIC suggested that this was optimized at smaller values of  $\sigma^2$  nearer 0.001. However, there was little practical difference ( $<5$ ) in the DIC derived from these smaller values for  $\sigma^2$  compared to the DIC derived from the optimal values for  $\sigma^2$  suggested through the optimization of  $P1$  and  $P2$  above. In fact, assessments of  $P1$  and  $P2$ , based only on match results up to earlier rounds during the 2003/2004 season, suggested that the optimum predictive performance at that stage is actually associated with slightly larger values of  $\sigma^2$ , above 0.004, and that the optimum value for  $\sigma^2$  reduces as the season progresses as more match results data are accumulated.

The choice of the common prior variance,  $\sigma_0^2$ , was also assessed in a similar way for various fixed values of  $\sigma^2$ . This suggested that the optimal choice of  $\sigma_0^2$  was in a broad range of between 0.01 and 0.02, with very little difference in model predictive performance over this range.

As an alternative, to having  $\sigma^2$  as a known constant in the model, it is now possible here to utilize the information above in order to formulate an informative prior for  $\sigma^2$ , with a view to retaining  $\sigma^2$  as a parameter in the model. Figure 2 suggests that a suitable prior for  $\sigma^2$  should cover a range of values between 0 and 0.02 and have a mean of 0.004. Therefore, a  $\text{Ga}(2,500)$  would seem to be a suitable informative prior in this case. However, the observation made above, that  $\sigma^2$  should be larger during the earlier part of the season, would support the use of a prior with a slightly larger mean and perhaps an increased range. In this case, a  $\text{Ga}(2,250)$  prior may be more appropriate. In fact, the choice of prior had little difference on either  $P1$ , which was 0.3974 for the  $\text{Ga}(2,250)$  prior and 0.3972 for the  $\text{Ga}(2,500)$  prior, or  $P2$ , which was 0.5453 for the  $\text{Ga}(2,250)$  and 0.5448 for the  $\text{Ga}(2,500)$  prior.

The convergence properties of the sampled values of  $\sigma^2$ , when retained as a parameter in the model, were not as good as those observed with all the remaining parameters. An example of the sample trace for this parameter is given in Fig. 3, which shows the sampled values using a  $\text{Ga}(2,250)$  prior and match results from the first three rounds of the 2003/2004 season. The sampled values exhibit a fair degree of snaking, which suggests that the sampler is not moving as freely around the parameter space as would be preferred. This suggests that a greater number of iterations than the previously reported minimum of 5,000 were required. Running the sampler for a total of 10,000 iterations, again with the first 2,500 being discarded as a burn-in, provided an MC error reported by WinBUGS which is close to the level of 5% (of the sampled standard deviation), which is normally considered as providing an acceptable sample. Therefore, the models that include  $\sigma^2$  as a parameter in the model were undertaken using 10,000 iterations.

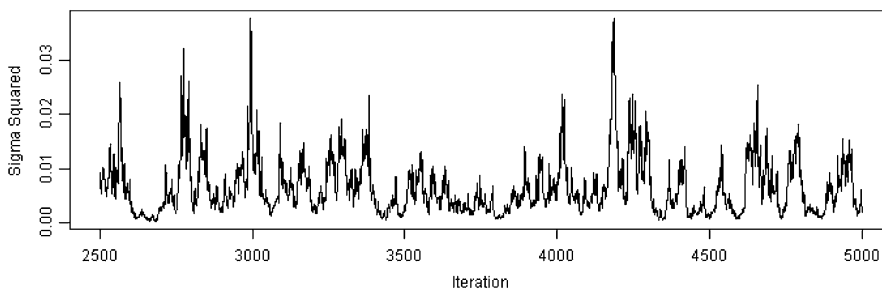


FIG. 3. Trace plots for  $\sigma^2$  as at round 3 for 2003/2004.



TABLE 1 *Example predictive probabilities for matches played in round 12 of the 2003/2004 season*

Fixture	Dynamic model			Non-dynamic model			Outcome
	$P(H)$	$P(D)$	$P(A)$	$P(H)$	$P(D)$	$P(A)$	
Celtic versus Dunfermline	0.87	0.09	0.04	0.85	0.10	0.05	Home win
Dundee versus Motherwell	0.43	0.25	0.32	0.53	0.23	0.24	Away win
Hibernian versus Dundee United	0.56	0.24	0.20	0.59	0.24	0.18	Draw
Livingston versus Partick	0.57	0.25	0.19	0.52	0.26	0.22	Home win
Aberdeen versus Hearts	0.22	0.21	0.57	0.26	0.25	0.48	Away win
Kilmarnock versus Rangers	0.08	0.13	0.79	0.09	0.15	0.76	Away win

### 3.2 Model assessment and parameter estimates

Recall that the measure  $P1$  represents the geometric mean of the one-match ahead predictive probabilities for the match outcomes that were actually observed. An example of these probabilities, which can easily be obtained from the model, are illustrated in Table 1. This lists the one-match ahead predictive probabilities of a home win,  $P(H)$ , draw,  $P(D)$ , and away win,  $P(A)$ , along with the eventual outcome, for each match that was to be played in Round 12 of the 2003/2004 season. These are predictive probabilities in the sense that they were derived by fitting the model only to match results data available up to and including Round 11 of that season. The predictive probabilities for both the dynamic and non-dynamic models are shown for comparison. Note that the results for the dynamic model presented in this section are based on the use of a  $\text{Ga}(2,250)$  prior for  $\sigma^2$ , with  $\sigma_0^2 = 0.02$ .

For the particular round of matches shown in Table 1, the probabilities contributing to  $P1$  from this round for the dynamic model are 0.87 (Celtic home win), 0.32 (Motherwell away win), 0.24 (Hibernian versus Dundee United draw), 0.57 (Livingston home win), 0.57 (Hearts away win) and 0.79 (Rangers away win). Note that each of these ‘observed outcome’ probabilities are higher than the corresponding probabilities from the non-dynamic model for this particular rounds of matches. However, note that this is not always the case, but the aim is to maximize the overall average observed outcome probabilities as measured by  $P1$ .

An objective means of assessing the relative predictive performance of the dynamic and non-dynamic models over time, as additional match results data are accumulated, is available through the use of cumulative forms of the predictive measures  $P1$  and  $P2$ , specified here as  $P1(t)$  and  $P2(t)$  as follows:

$$P1(t) = \exp \left\{ \frac{1}{N(t)} \sum_{k=1}^{N(t)} \log e[P(O_k)] \right\}, \quad k = 1, 2, \dots, N(t), \quad (21)$$

$$P2(t) = \frac{1}{N(t)} \sum_{k=1}^N \left\{ [1 - P(O_k)]^2 + P(NO_{1k})^2 + P(NO_{2k})^2 \right\}, \quad k = 1, 2, \dots, N(t), \quad (22)$$

where  $N(t)$  is the number of matches played during a particular season up to and including round  $t$ .

Figures 4 and 5 display plots of  $P1(t)$  and  $P2(t)$ , respectively, against  $t$  for each of the three seasons: 2003/2004, 2004/2005 and 2005/2006. These are plotted from Round 9 onwards to allow the cumulative values of  $P1(t)$  and  $P2(t)$  to include at least 2 months data. Figures 4 and 5 clearly illustrate that the dynamic model provides superior predictive performance, compared to the non-dynamic model, as measured by higher levels of  $P1(t)$ , and lower levels of  $P2(t)$  and that this is consistent across all three seasons considered. The  $\text{Ga}(2,250)$  prior employed for the evolution variance,  $\sigma^2$ , would therefore

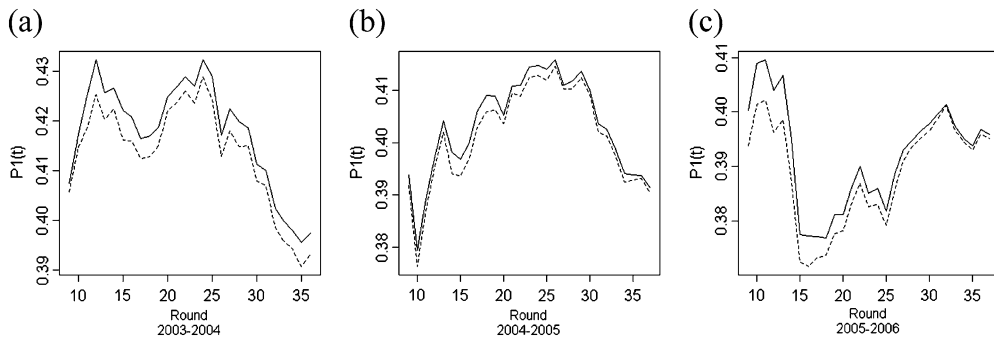


FIG. 4. Plots of  $P1(t)$  for the dynamic (—) and non-dynamic (---) models for 2003/2004 (a), 2004/2005 (b) and 2005/2006 (c).

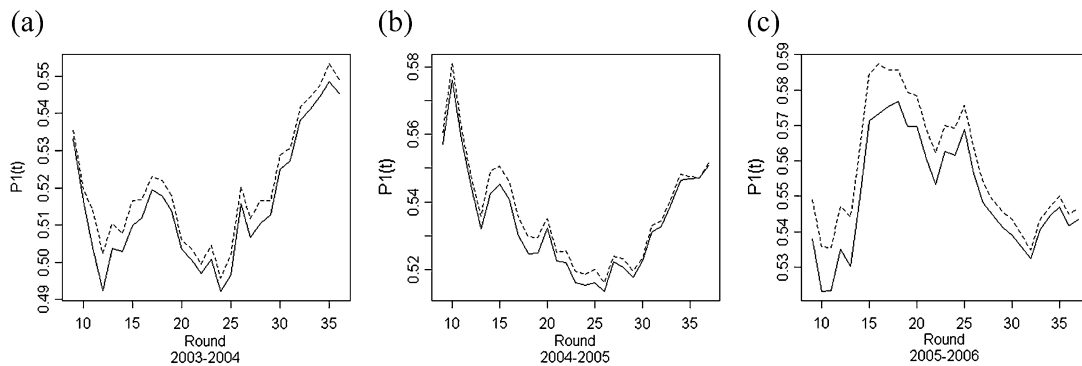


FIG. 5. Plots of  $P2(t)$  for the dynamic (—) and non-dynamic (---) models for 2003/2004 (a), 2004/2005 (b) and 2005/2006 (c).

appear to be a suitable choice for the analysis of other seasons' data in the SPL and for real-time deployment of the model in the SPL. This could also provide a suitable prior for other football leagues. Note however, as reported previously, the predictive performance of the model, based on  $P1(t)$  and  $P2(t)$ , was also found to differ very little between the two alternative priors for  $\sigma^2$ ,  $\text{Ga}(2,250)$  and  $\text{Ga}(2,500)$ , such that either choice of prior provides superior predictive performance over the non-dynamic model.

The manner in which the updated posterior estimates of the evolution variance ( $\sigma^2$ ) are 'updated' as additional match results data become available, is of interest since determination of this parameter is the hardest and most crucial aspect of the work described in this paper. Figure 6 illustrates how these posterior estimates are updated over time as additional match results data are accumulated for both choice of priors over each season. Not surprisingly, in each season, the estimates for  $\sigma^2$  start at around 0.008 with the  $\text{Ga}(2,250)$  prior, and 0.004 with the  $\text{Ga}(2,500)$ , which simply reflect the respective means of these prior distributions when little match results data are available. It is interesting to note that Fig. 6(b and c) for 2004/2005 and 2005/2006, respectively, indicate that the estimate of  $\sigma^2$  does indeed initially remain high at the start of these seasons, which allows for greater adaption in the attack and defence parameters. The fact that this feature is not evident in Fig. 6(a) for 2003/2004, is most likely due to the fact that better prior information was available for that season since, no team was relegated from, or promoted to, the SPL at the end of the 2002/2003 season as discussed earlier. Figure 6(a) does however indicate that the estimate of  $\sigma^2$  actually increases slightly during Rounds 10–14, which suggests during this period, that several teams' results went against the pattern of results in previous rounds. The ability for the model to adjust the estimate of  $\sigma^2$  as required to allow more or less adaption in

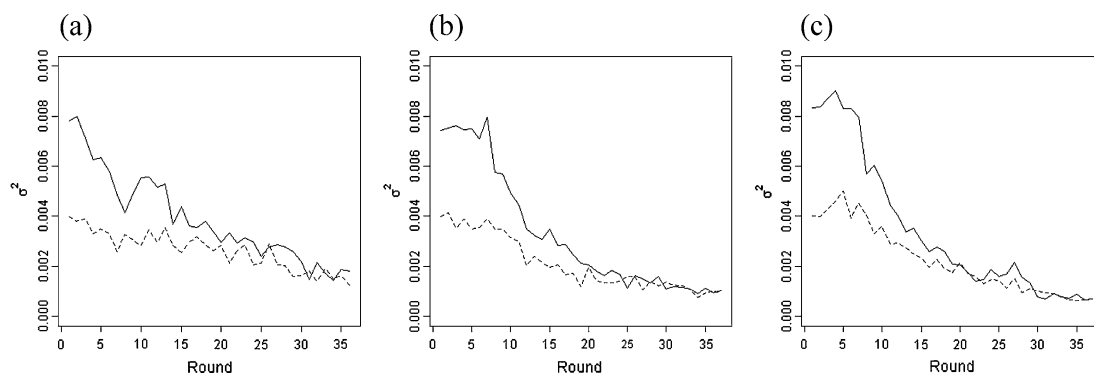


FIG. 6. Plots of posterior estimates for  $\sigma^2$  using Ga(2,250) (—) and Ga(2,500) (---) Priors for 2003/2004 (a), 2004/2005 (b) and 2005/2006 (c).

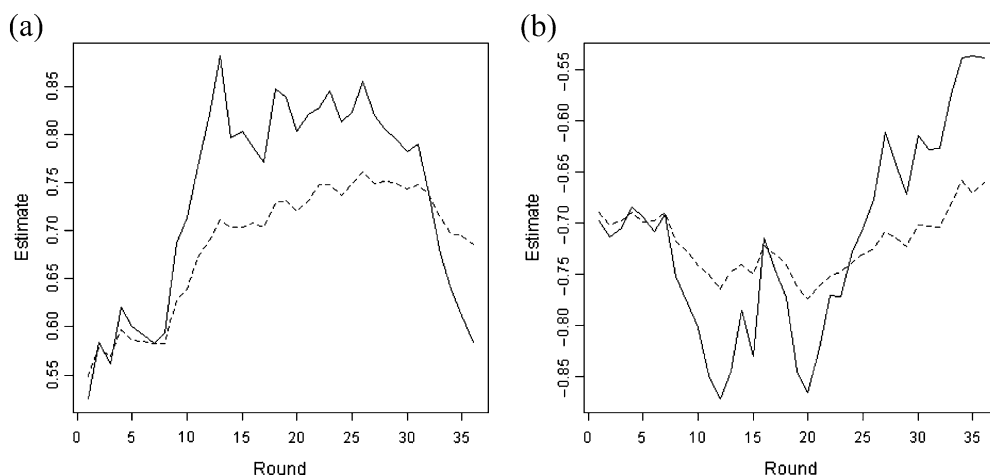


FIG. 7. Posterior estimates of attack (a) and defence (b) parameters for Celtic in 2003/2004. Dynamic (—) model and non-dynamic (---) model.

the parameter estimates clearly represents a key advantage of the dynamic modelling approach outlined in this paper. Figure 6 also indicates that the estimate of  $\sigma^2$  eventually reduces down to around 0.001 in each season, which is consistent with the observations made previously, where the DIC was minimized with smaller values for  $\sigma^2$  at around this value.

The effect on the estimates of the attack and defence parameters by including a dynamic component in the model is illustrated in Fig. 7. This shows the evolution of the attack and defence parameters for Celtic during the 2003/2004 season, using both the dynamic and non-dynamic models. The updated estimates shown are those that were estimated as each new round of match results data became available. Hence, they do not reflect the view of the how the teams' abilities had changed if looking back at the end of the season in which case the non-dynamic model would be reflected by a constant horizontal line since that model assumes the parameter estimates remain constant through time.

The case of Celtic in this particular season is interesting since they scored a record 105 goals that season, winning the SPL title by Round 32. However, having secured the title after 32 rounds of matches and scoring 98 goals, they subsequently managed to score just seven goals in their final six matches.

Figure 7(a) illustrates how the dynamic model captures this severe change in attacking ability, by adapting the estimates of the attack parameter for Celtic after Round 32, much more quickly than the non-dynamic model. In addition, in that same season, Celtic had conceded just seven goals during the first 20 rounds of matches, but subsequently conceded 18 goals in their remaining 18 matches. Again, this significant change in defensive ability is reflected in Fig. 7(b), which shows how the dynamic model again adapts much more quickly to this change, compared to the non-dynamic model.

#### 4. Comparison with other approaches

The dynamic model described here has similarities with that discussed in Rue & Salvesen (2000), which considers an approach to the evolution component in the model based on continuous time, instead of the discrete time approach adopted here. Those authors also employ several other modifications to their Poisson model that were considered appropriate for data from the English leagues. However, our model has been applied to data the SPL for which there is no evidence that similar additional modifications to the Poisson model are required.

The approach described in Rue & Salvesen (2000) also assumes that the evolution variance,  $\sigma^2$ , and the prior variance,  $\sigma_0^2$ , are common to the attack and defence abilities for all teams. Their specification of the evolution variance between two continuous time points,  $t$  and  $s$  ( $s > t$ ), is equivalent to

$$\sigma^2 = \sigma_0^2 \left( \frac{s - t}{\nu} \right). \quad (23)$$

The parameter,  $\nu$ , in (23) is used in Rue & Salvesen (2000) to control the size of the evolution variance. Based on match results data from English football, the values of  $\nu$  and  $\sigma_0^2$ , which optimize the predictive performance of the model, are reported in Rue & Salvesen (2000) to be 100 and 1/37, respectively. Note that in our model, the optimal value for  $\sigma_0^2$  is in the range 0.01–0.02 (1/50–1/100), although it should be noted that Rue & Salvesen (2000) uses ‘less accurate’ prior means for the log-attack and log-defence parameters. Their prior means were equally spaced on the interval from  $-0.2$  to  $+0.2$  (based on their prior rankings of the teams), whereas the work presented here uses ‘more accurate’ prior means in the sense that they are derived from applying a model to previous seasons’ data.

In a typical season in English or indeed Scottish football, the time between successive matches for a team is most commonly either 3–4 days, 1 week or 2 weeks. Therefore, the values of  $s - t$ , given by (23), are typically restricted to values of 3.5, 7 or 14. Hence, using values of 100 and 1/37 for  $\nu$  and  $\sigma_0^2$ , respectively, the value of  $\sigma^2$  in (23) will typically be restricted to values close to either 0.001, 0.002 or 0.004. Interestingly, these fall in the same range covered by the model-fitted estimates for  $\sigma^2$ , in the context of Scottish football, presented earlier in Fig. 6. It is therefore no surprise to find that our discrete time approach presented in this paper, typically provides similar one-match ahead predictive probabilities of future match outcomes, as those derived from the continuous time approach. This similarity is illustrated clearly in Fig. 8, which plots  $P1(t)$  and  $P2(t)$  for the 2003/2004 season in the SPL, using both our discrete time approach and the continuous time approach given by (23) above. In this analysis,  $\nu$  and  $\sigma_0^2$  were taken to be 100 and 0.02, respectively, and the prior means used were those given in the Appendix to this paper.

It is worth highlighting, however, that the discrete time approach is a required aspect of the model if parameter constraints are to be defined since these need to be assumed to hold at each time point (e.g. round) during the season. In fact, the continuous time approach of Rue & Salvesen (2000) is facilitated by removing the need for any parameter constraints, through the estimation of the parameters  $\gamma_H$  and  $\gamma_A$  using historical data rather than estimating these using the latest information by retaining them as parameters in the model. Rue & Salvesen (2000) actually estimate  $\gamma_H$  and  $\gamma_A$  using the (logarithm)

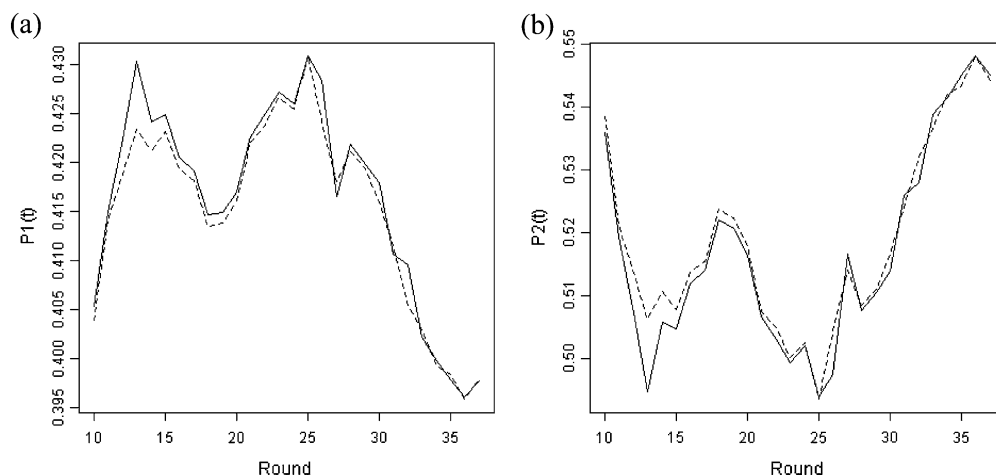


FIG. 8. Plots of  $P1(t)$  (a) and  $P2(t)$  (b) for 2003/2004. Discrete time (—) and continuous time (---).

of the mean goals scored home and away in the previous four seasons. This presents several potential problems, the most obvious of which is that the continuous time model will perform less well if the average rate at which goals are being scored at home and/or away during the current season, differs markedly from that experienced in the past. In contrast, the discrete time model presented in this paper estimates  $\gamma_H$  and  $\gamma_A$  using the latest data and hence presents an advantage in that respect over the continuous time approach.

The elicitation of effective priors for the attack and defence parameters is also more difficult in the continuous time model since in that model, the attack and defence parameters are interpreted relative to historical estimates for  $\gamma_H$  and  $\gamma_A$  rather than the current estimates. The attack and defence parameters in our discrete model are always interpreted relative to the current level of goals scored. Hence, in our discrete time model, a team performing at a similar level of attacking or defensive ability from one season to the next will have similar attack and defence parameters carried over from one season to the next. However, this is not necessarily the case in the continuous time model.

The final advantage that our discrete time approach offers, which is the most significant, relates to the fact that this allows the evolution variance,  $\sigma^2$ , to be estimated as a parameter in the model. This is perhaps most important in situations where the prior means for the attack and defence parameters have been mis-specified and do not adequately reflect the true differences in abilities between the teams. As an example, Fig. 9 plots  $P1(t)$  and  $P2(t)$ , again for the 2003/2004 season in the SPL, using both our discrete time approach and the continuous time approach, but this time using less accurate prior means. These prior means were based on values that were equally spaced on the interval from  $-0.2$  to  $+0.2$  (using the prior rankings of the teams implied by the information provided in the Appendix). The values of  $\nu$  and  $\sigma_0^2$  were again taken to be 100 and 0.02, respectively. In the case of the discrete time model, the fact that the estimated evolution variance is estimated as a parameter in the model with a  $Ga(2,250)$  prior, allows this to remain at a higher level when greater adaption is required. In fact, the evolution variance in this model remains high, at values around 0.008 or above, until Round 16, which as Fig. 9 illustrates, it allows the parameter estimates in the model to adapt much more quickly to values that are more representative of teams relative abilities. In the continuous time model, however, recall that the evolution variance is effectively fixed at values near to either 0.001, 0.002 or 0.004, which causes the

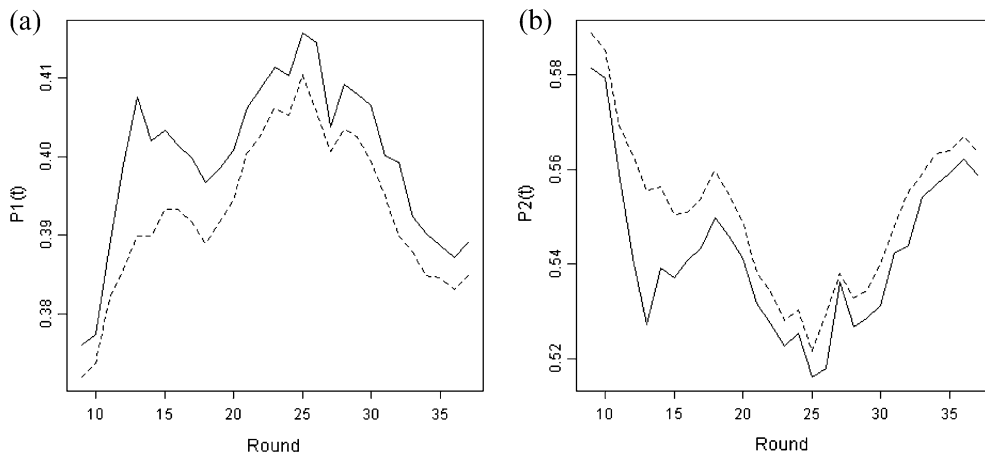


FIG. 9. Plots of  $P1(t)$  (a) and  $P2(t)$  (b) for 2003/2004. Using less accurate prior means. Discrete time (—) and continuous time (---).

parameter estimates to adapt much more slowly, which is illustrated in Fig. 9. The discrete time model does recover towards the end of the season, but it is very slow to do so in this situation.

## 5. Discussion and proposed further work

This paper has outlined a practical approach to implementing a dynamic model for estimating outcome probabilities for future football matches and provides evidence that the dynamic model offers improved predictive performance over the non-dynamic form of the model. The choice of evolution variance,  $\sigma^2$ , was shown to be crucial in terms of optimizing the predictive performance of the model, and an approach that allows this to be estimated as a parameter in the model has been described. A  $\text{Ga}(2,250)$  prior for this parameter seemed to provide optimal predictive performance in the model, which allows the evolution variance to take higher values at the start of the season when greatest adaption in the parameter estimates is required. The model is based on a discrete time approach that has been shown to offer advantages over a continuous time approach. Further development of the model presented here to allow the time development of the evolution variance to be modelled in a more direct way would seem a natural additional consideration for future work. However, in the meantime, the choice of prior here would appear to offer a suitable choice of prior for other seasons within the SPL and perhaps even other football leagues.

The approach used with regard to deriving suitable priors for the attack and defence parameters for the promoted sides was rather simplistic. Further work is planned to optimize this aspect of the modelling process before true comparisons can be made with other providers of match outcome probabilities.

Finally, both the dynamic and non-dynamic versions of the model described here have been observed to significantly over-estimate the probability of a 0–0 draw in the SPL. This is quite a different situation to that observed in many other European football leagues, where the models have been found to under-estimate the probability of a 0–0 draw, particularly in the Italian Serie A. An approach to dealing with this problem that seems to be particular to the SPL is currently being investigated.

## Acknowledgement

The author is indebted to Dr Karen Vines and Prof. Kevin McConway of the Open University, UK, for their support and advice provided with preparing this paper and during his ongoing Ph.D. research in this area.

## REFERENCES

- CROWDER, M., DIXON, M., LEDFORD, A. & ROBINSON, M. (2002) Dynamic modelling and prediction of English Football League matches for betting. *Statistician*, **51**, 157–168.
- DIXON, M. J. & COLES, S. G. (1997) Modelling association football scores and inefficiencies in the football betting market. *Appl. Stat.*, **46**, 265–280.
- KARLIS, D. & NTZOUFRAS, I. (2003) Analysis of sports data by using bivariate Poisson models. *Statistician*, **52**, 381–393.
- KNORR-HELD, L. (2000) Dynamic rating of sports teams. *Statistician*, **49**, 261–276.
- LEE, A. J. (1998) Modelling scores in the Premier League: is Manchester United really the best? *Chance*, **10**, 15–19.
- MAHER, M. J. (1982) Modelling association football scores. *Stat. Neerl.*, **36**, 109–118.
- OWEN, A. (2009) Dynamic Bayesian forecasting models of football match outcomes. *Proceedings of the Second International Conference on Mathematics in Sport* (D. F. Percy & P. A. Scarf eds). Groningen, The Netherlands: Institute of Mathematics and Applications, pp. 135–140.
- RUE, R. & SALVESEN, Ø. (2000) Prediction and retrospective analysis of soccer matches in a league. *Statistician*, **49**, 399–418.
- WEST, M. & HARRISON, J. (1997) *Bayesian Forecasting and Dynamic Models*. New York: Springer.

## Appendix

Values for  $m_\alpha$ ,  $m_\beta$ ,  $g_H$ ,  $h_H$ ,  $g_A$  and  $h_A$  used in the priors for the analysis of each of the three seasons from 2003/2004 to 2005/2006

	2003/2004		2004/2005		2005/2006	
	$m_\alpha$	$m_\beta$	$m_\alpha$	$m_\beta$	$m_\alpha$	$m_\beta$
Aberdeen	−0.279	0.018	−0.273	0.127	−0.169	−0.062
Celtic	0.591	−0.685	0.667	−0.663	0.590	−0.495
Dundee	−0.026	0.083	−0.042	0.109	N/A	N/A
Dundee United	−0.416	0.241	−0.232	0.170	−0.134	0.279
Dunfermline	0.073	0.256	−0.009	0.113	−0.198	0.193
Falkirk	N/A	N/A	N/A	N/A	−0.155	0.178
Hearts	0.088	−0.081	0.088	−0.179	−0.005	−0.183
Hibernian	0.060	0.218	−0.072	0.196	0.153	0.178
Inverness C.T.	N/A	N/A	−0.037	0.234	−0.257	0.083
Kilmarnock	−0.093	−0.005	−0.079	0.143	−0.010	0.195
Livingston	−0.098	0.170	−0.155	0.096	−0.199	0.202
Motherwell	−0.157	0.300	−0.321	0.191	−0.080	0.037
Partick Thistle	−0.374	0.084	N/A	N/A	N/A	N/A
Rangers	0.630	−0.599	0.466	−0.536	0.464	−0.605
$g_H$	216		216		199	
$h_H$	147		147		141	
$g_A$	131		120		111	
$h_A$	115		110		105	