Dynamic bayesian forecasting models of football match outcomes

Dynamic Bayesian forecasting models of football match outcomes

Alun Owen*

*Faculty of Engineering and Computing, Coventry University Coventry, CV1 5FB, aa5845@coventry.ac.uk

Abstract. Dynamic Generalized Linear Models (DGLMs) are essentially generalised linear models with parameters that are stochastic. They are Bayesian in flavour and are particularly suited to forecasting applications. This paper outlines a practical implementation of a Poisson DGLM model that can easily be deployed using the freely available software WinBUGS. Using match results data from the Scottish Premier League (SPL) between 2003/2004 to 2005/2006, the DGLM approach is shown to provide more improved predictive probabilities of future match outcomes, compared to the non-dynamic form of the model.

1. Introduction

Statistical modelling of association football match data is often of interest with regard to either developing team rankings, or deriving predictive probabilities of future match outcomes, which are typically in terms of a home win, draw or away win. Much of the published literature in this respect has considered models from a classical standpoint, typically making use of the Generalized Linear Modelling (GLM) framework and using maximum likelihood methods for parameter estimation. One problem with this approach is that the parameters in the model are assumed to remain constant over time, which would seem unrealistic given the potential variable nature of individual team's performance over time. This paper therefore presents an approach based on the use of the Dynamic Generalized Linear Modelling (DGLM) framework, described in West and Harrison (1997), which allows some or all of the parameters in the model to time dependent. DGLMs are well suited to forecasting applications and have very much a Bayesian flavour, typically requiring the use of Bayesian approaches to facilitate parameter estimation.

A Poisson DGLM model is applied here in a Bayesian framework, to match results data from the Scottish Premier League (SPL). Parameter estimates and predictive probabilities of future match outcomes are derived through MCMC methods, using the freely available software WinBUGS (see http://www.mrc-bsu.cam.ac.uk/bugs). Section 2 describes the structure of the model, which includes consideration of an interesting problem of dynamically modelling parameters which are subject to constraints. An assessment of the model's predictive performance is presented in Section 3, with some comments and a discussion of proposed future work outlined in Section 4.

2 A Dynamic Generalized Linear Model

2.1 Model specification

The model developed here is based on a (non-dynamic) model, originally presented in Maher (1982) and also considered by a number of other authors including Dixon and Coles (1997) and Karlis and Ntzoufras (2003). However, the model is extended to the dynamic case and a slightly different parameterisation is used, 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:

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

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

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

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

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:

$$\sum_{i=1}^{n} \alpha_{i,t} = 0, \tag{5}$$

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

The attack and defence parameters therefore represent the attacking and defensive strengths, relative to an average team which have average attack and defence parameters of 0. The parameters γ_H and γ_A therefore reflect the underlying (natural logarithm of) overall average scoring rates at home and away respectively and are assumed to remain constant over time.

This parametrisation differs from that used by some of the the authors mentioned previously, where typically only a single identifiability constraint was required. The advantage gained by the parameterisation used here is two-fold; firstly it provides improvements to the convergence properties of the defence parameters where an MCMC approach is used, which is illustrated later in 2.3, and secondly the model has the attractive symmetrical property in that the attack and defence parameters are treated equally in the model.

The terms specified in (1) to (4) form the *observation* component of the model, and to fully specify this as a dynamic model, an *evolution* component is 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:

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

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

where the parameter τ represents the *evolution precision* (reciprocal of the variance). For simplicity, the evolution precision is assumed here to remain constant over time, and to be common to all teams and common to both the attack and defence parameters. It is quite straight forward to extend this to the case where theses assumptions are relaxed.

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

$$\alpha_{i,0} \sim N(m_{\alpha_i}, \tau_0^{-1}), \quad (9)$$

$$\beta_{i,0} \sim N(m_{\beta_i}, \tau_0^{-1}), \quad (10)$$

$$\gamma_H^{'} \sim Ga(g_H, h_H), \quad \gamma_A^{'} \sim Ga(g_A, h_H), \quad (11)$$

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_α and m_β are known constants representing prior means for these baseline attack and defence strengths. The parameter τ_0 represents the common *prior precision*, 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 and Harrison (1997) and Knorr-Held (2000) are used to overcome this problem. Firstly for the attack parameters, to make the notation easier, we define:

$$\boldsymbol{\alpha}_t = [\alpha_{1,t}, \alpha_{2,t}, ..., \alpha_{n,t}]^T , \quad (12)$$

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

The evolution of the attack parameters (7) can then be expressed as $\boldsymbol{\alpha}_t \sim N(\boldsymbol{\alpha}_{t-1}, \mathbf{W})$, where $\mathbf{W}$ is a diagonal evolution variance matrix with entries τ^{-1}. Similarly, the initial priors on the attack parameters (9) can be expressed as $\boldsymbol{\alpha}_0 \sim N(\mathbf{m}_\alpha, \mathbf{W}_0)$, where $\mathbf{W}_0$ is a diagonal prior variance matrix with entries τ_0^{-1}. 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 $\mathbf{1}_n^T = [1,1,...,1]$, and if the evolution variance matrices, $\mathbf{W}$ and $\mathbf{W}_0$, are modified to variance-covariance matrices $\mathbf{R}$ and $\mathbf{R}_0$, resepectively, as follows:

$$\mathbf{R} = \frac{n}{(n-1)\tau}(\mathbf{I}_n - \frac{1}{n}\mathbf{1}_n\mathbf{1}_n^T), \tag{14}$$

$$\mathbf{R}_0 = \frac{n}{(n-1)\tau_0}(\mathbf{I}_n - \frac{1}{n}\mathbf{1}_n\mathbf{1}_n^T), \tag{15}$$

where $\mathbf{I}_n$ is the $n \times n$ identity matrix. Note that the multiplying factor $n/(n\text{-}1)$ is incorporated so that the variances on the diagonals reflect the variances (or precisions) 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 conditonal 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 $\mathbf{c}_t = [\theta_{1,t}, \theta_{2,t}, ..., \theta_{n\text{-}1,t}]^T$, from a multivariate normal distribution with zero mean and variance-covariance matrix $\mathbf{S}_t$ given by:

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

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

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

it can be shown that the resulting values of $\boldsymbol{\alpha}_t = \boldsymbol{\alpha}_{t\text{-}1} + \mathbf{u}_t$ 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 $\mathbf{c}_0 = [\theta_{1,0}, \theta_{2,0}, ..., \theta_{n\text{-}1,0}]^T$, from a normal distribution with zero mean and variance-covariance matrix $\mathbf{S}_0$ given by:

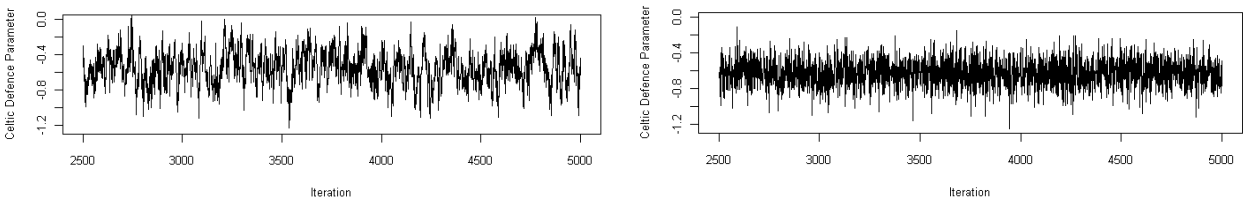
$$\mathbf{S}_0 = \frac{n}{(n-1)\tau_0}(\mathbf{I}_{n-1} + \mathbf{1}_{n-1}\mathbf{1}_{n-1}^T) \tag{18}$$

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

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

2.3 Model implementation and optimisation

The model was deployed retrospectively, on a round by round basis, using match results data from the SPL over each season from 2003/2004 to 2005/2006. The sampled values displayed very good mixing behaviour and running the sampler for a minimum of 5,000 iterations, with the first 2,500 iterations being used as a burn in, was assessed as being adequate for estimation purposes. Parameter estimates were thus derived from the last 2,500 iterations. One down-side to incorporating the additional identifiability constraint on the defence parameters (6), and hence the additional parameter γ_A in the model, was a significant increase in the time required to run each sample of 5,000 iterations. However, there are improvements in the mixing behaviour of the sampled values for the defence parameters, to be derived by incorporating this additional constraint. This is illustrated in the example sample traces shown in Figure 1, which relate to the posterior estimates of the latest defence parameter for Celtic after three rounds of matches had been played during the 2003/2004 season. Figure 1(a) displays the sampled values where the additonal constraint is not included, which exhibits some degree of snaking. This is indicative of significant autocorrelation between the sampled values, which is less than satisfactory when deriving parameter estimates using MCMC methodology. This can be contrasted with Figure 1 (b), where the additonal constraint is included, which indicates improved sampling behaviour and is much more satisfactory.



(a) (b)

Figure 1: Trace Plots for Celtic Defence Parameter at Round 3 in 2003/2004
Without Additional Identifiability Constraint (a) and With Additional Identifiability Constraint (b)

Values for the known constants m_α, m_β, g_H, h_H, g_A and h_A , specified in the priors in (9) to (11), were derived from fitting the non-dynamic model to the previous season's full set of match results. The values for these constants used in the analysis of the 2003/2004 season, are shown in the appendix. In the SPL, one team is relegated and replaced by a team promoted from the Scottish League Division 1. As a simplistic approach, the priors derived for 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. Research into optimising this aspect of the modelling process is ongoing.

The evolution precision parameter, τ, can also be pre-specified as a known constant or kept as a parameter in the model to be estimated. However, if estimated as a parameter in the model, the resulting posterior estimates were very sensitive to the choice of prior. Therefore τ was also pre-specified as a known constant. Since choice of value for this parameter is crucial in terms of modelling the stochastic changes in the attack and defence parameters, this was determined primarily by optimising the model's short term predictive performance. One commonly used measure of short-term predictive performance in these types of models, is that defined as P1, which is based on the N matches played over one complete season as follows:

$$\mathrm{P1} = \exp\left\{\frac{1}{N}\sum_{k=1}^{N}\log e\left[P(O_k)\right]\right\} \quad k = 1,2,\ldots,N, \tag{19}$$

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, for example by Knorr-Held (2000), is that defined as P2 as follows:

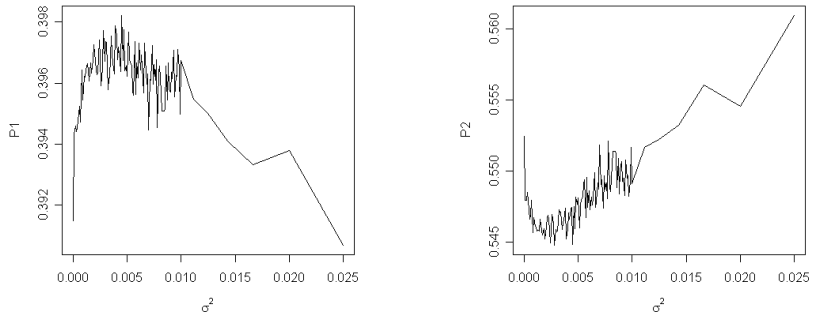
$$\mathrm{P2} = \frac{1}{N}\sum_{k=1}^{N}\left\{\left[1 - P(O_k)\right]^2 + P(NO_{1k})^2 + P(NO_{2k})^2\right\} \quad k = 1,2,\ldots,N, \tag{20}$$

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. The effect of different values of τ on the overall model fit was also investigated through use of the Deviance Information Criterion (DIC). This 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. The effect of the choice of the prior precision parameter, τ_0, on the predictive performance and model fit measures was also investigated.

3 Results

The effect of the choice of evolution precision, τ, on the short-term predictive performance of the model, is illustrated in Figure 2 below. This plots the resulting values of P1 and P2 for various values of the evolution variance σ^2 (=1/ τ), based on the full set of one-match ahead predictive probabilities for the 2003/2004 season in the SPL. These plots include the values of P1 and P2 derived from the non-dynamic model, which is equivalent to $\sigma^2 = 0$. Figure 2(a) suggests that P1 is actually optimised for values of σ^2 near 0.004 (τ =

250), whilst Figure 2(b) suggests P2 is optimised for slightly smaller values of σ^2 near 0.0025 ($\tau = 400$). Note that values for σ^2 in the range from 0 to 0.01 were investigated at intervals of 0.0001, in order to identify the optimal region, whereas values for σ^2 were investigated at much less frequent intervals as σ^2 increased above 0.01, in order to verify the continuing behaviour of P1 and P2.



(a) (b)

Figure 2: P1 (a) and P2 (b) versus Evolution Precision (σ^2) for 2003/2004 Season

The DIC, however, was minimised for smaller values of σ^2 near 0.001 ($\tau = 1000$), although there was little practical difference in the DIC between this optimum and the optimal values for σ^2 suggested by the assessment of P1 and P2 above. Since our primary interest is in terms of short-term prediction, the optimum choice for σ^2 was taken be 0.004 ($\tau = 250$) and used throught the remainder of the analyses presented here.

A similar assessment of the optimum choice of common prior precision τ_0, suggested this was in a broad range of between 50-100, with very little difference over this range.

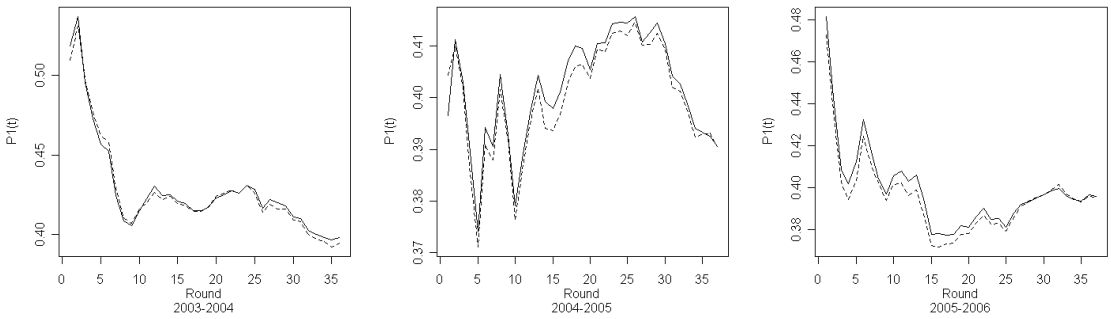
As a way of assessing the relative predictive performance of the dynamic and non-dynamic models over time as each season progresses, cumulative forms of the predictive measures P1 and P2, are considered and specified here as P1(t) and P2(t) as follows:

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

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

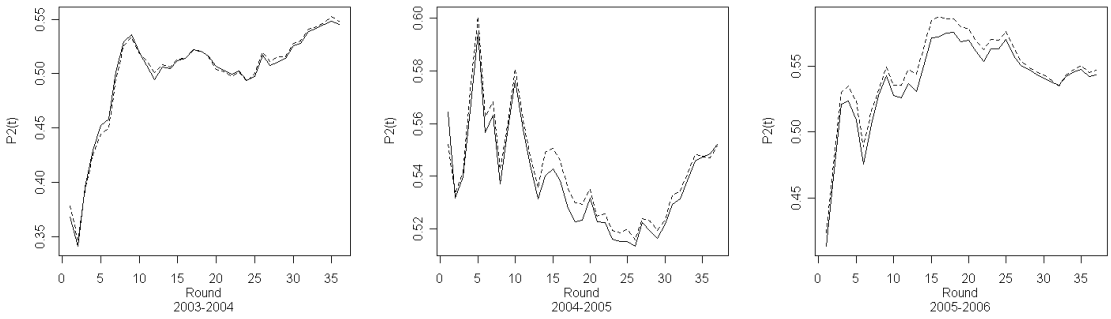
where $N(t)$ is the number of matches played during a particular season up to and including round t.

Figures 3 and 4 display plots of P1(t) and P2(t) respectively, against t, for each of the three seasons 2003/2004, 2004/2005 and 2005/2006. The plots show the results for both the non-dynamic model, and the dynamic model with $\tau = 250$ and $\tau_0 = 50$. These suggest that the dynamic model is almost always at least as competitive as the non-dynamic model, but regularly displays a superior predictive performance as measured by higher levels of P1(t), and lower levels of P2(t).



(a) (b) (c)

Figure 3: Plots of P1(t) for the Dynamic (—) and Non-dynamic (---) models for 2003/2004(a), 2004/2005(b) and 2005/2006(c) Seasons



(a) (b) (c)

Figure 4: Plots of P2(*t*) for the Dynamic (—) and Non-dynamic (---) models for 2003/2004(a), 2004/2005(b) and 2005/2006(c) Seasons

4 Discussion and proposed further work

This paper provides evidence that the dynamic model offers improved predictive performance over the non-dynamic model. However, the approach used to deriving suitable priors was rather simplistic, and so further research is required to optimise this aspect of the modelling process, before comparisons can be made with other providers of match outcome probabilities. The author notes that a more general version of the dynamic model considered was discussed in Crowder et. al. (2002). However, this differs from the work presented here since those authors derived parameter estimates via an approximation method, and the aim here was to describe an implementation of a dynamic model that can be deployed easily using readily available software. 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. An approach to dealing with this problem which makes use of so called 'hurdle' models is currently being investigated.

Acknowledgements

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

Appendix

Values for the known constants m_α, m_β, g_H, h_H, g_A and h_A used in the analysis for SPL 2003/2004.

	m_α	m_β
Aberdeen	-0.279	0.018
Celtic	0.591	-0.685
Dundee	-0.026	0.083
Dundee United	-0.416	0.241
Dunfermline	0.073	0.256
Hearts	0.088	-0.081

	m_α	m_β
Hibernian	0.060	0.218
Kilmarnock	-0.093	-0.005
Livingston	-0.098	0.170
Motherwell	-0.157	0.300
Partick Thistle	-0.374	0.084
Rangers	0.630	-0.599

g_H	216
h_H	147
g_A	131
h_A	115

References

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

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

Karlis D. and 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. *The Statistician* 49, 261–276.

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

West M. and Harrison J. (1997) *Bayesian Forecasting and Dynamic Models*, Springer.