Fast sequential Bayesian prediction of football using mixtures of state space models

Gareth Ridall,Lancaster University, UK,
Anthony Pettitt, QUT, Australia,
email: g.ridall@Lancaster.ac.uk

June 29, 2019

Fast sequential Bayesian prediction of football using mixtures of state space models

Gareth Ridall,Lancaster University, UK,
Anthony Pettitt, QUT, Australia,
email: g.ridall@Lancaster.ac.uk

June 29, 2019

Table of contents

Abstract

Abstract

- We look at the history of premier league football over the last 23 years using a mixture of state space models where the states are defensive and attacking form (The ability to score goals and restrict goals).
- We show that the seasons vary a lot in volatility. Some seasons are predictable with teams maintaining their form over the season. In other seasons, the form of the teams vary substantially throughout the season with some teams exhibiting sharp changes of style and form. Plots
- To model this we use a mixture of models to accommodate both sequences of games of predictable form and other sequences of games showing highly variable form. In using mixtures we arrive at a model with superior predictive properties.

Methodological Background

- West, Harrison and Migon (1985) introduce a class of dynamic generalised linear models where dynamic updates of the sufficient statistics can be made through the exploitation of conjugacy.
- We extend this methodology to models where the dynamic parameters do not have sufficient statistics but where the full conditional posteriors of each parameter are from known distributions.
- We show, using applications, how proxys for the sufficient statistics (which we call quasi-sufficient statistics) can be constructed for the purpose of sequential updating and marginalisation and prediction.
- Parameters that cannot be updated in this way are averaged over. These include discount factors which represent the rate of change of form and another parameter to model correlation and over-dispersion.

Introduction and background

Main ingredients of methodology

We exploit these ideas which simplify and speed up inference

1. Sequential Bayes with parameter discounting (forgetting the past).
2. Exploitation of conjugacy providing closed form expressions for posterior and cumulative evidence
3. Mixtures of models with differing discount factors (each weighted by its cumulative evidence).
4. Use of sufficient statistics or proxys for them to be used to define conditional posteriors
5. Use of only multiplicative models
6. Total absence of simulation based methods

The state space model

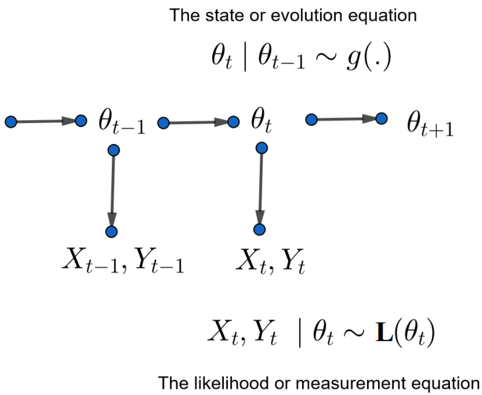

Figure: Sequential evolution of parameters of two conditionally independent Poisson distributions.

Predicting premier league football outcomes

Aim of the analysis

To develop a model or combination of models capable of dynamically predicting the premier league football outcomes over the last two decades, using only the final scores, date and the home ground advantages. In this way we hope to develop an exploratory tool to identify changes in form and style both within and between the seasons of the major clubs.

Results, dates of results and bookmaker odds of results of over 20 seasons of premier league football are available online at

http://www.football-data.co.uk/englandm.php

Literature

- Residual analysis, from stationary univariate Poisson models over each season, indicate low, predominantly positive correlation. In addition there is some over-dispersion.
- Dixon and Coles (1997) formulate correlation only between low scoring games.
- Karlis and Ntzoufras (2003), Crowder et. al (2002) and Koopman and (2013) use latent variables to explain positive correlations.
- Gamerman et al.,(2013) construct conjugate state space models for univariate models
- We suggest a shared gamma multiplicative random effect to explain both over-dispersion and positive correlations.

The likelihood and priors

The likelihood

Let $i \in \{1, 2, \ldots, 20\}$ denote the home team and $j \in \{1, 2, \ldots, 20\}$ denote the away team and let the games, in chronological order be labelled as $t = 1, \ldots, 380$.

- Let $\alpha_{i,t}$ be the attacking strength of team i at time t,
- Let $\beta_{j,t}$ be the defensive strength of team j at time t
- let γ_t be the common home ground advantage.

Then the home goals X_t and away goals Y_t at time t are conditionally independent Poisson distributions

$$\begin{pmatrix} X_t \\ Y_t \end{pmatrix} \mid \kappa \sim \mathrm{Poisson}\begin{pmatrix} \alpha_{i,t}\beta_{j,t}\gamma_t\epsilon_t \\ \alpha_{j,t}\beta_{i,t}\epsilon_t \end{pmatrix} \qquad \text{(Likelihood)}$$

where $\epsilon_t \sim \mathrm{Gamma}(\kappa, \kappa)$ is a shared random effect for game t.

The likelihood

The joint likelihood of data and random effects is

$$f(\mathbf{x},\mathbf{y},\boldsymbol{\epsilon} \mid \boldsymbol{\alpha},\boldsymbol{\beta},\gamma_t,\kappa) \propto \exp(-\sum_{t=1}^{380}\epsilon_t[\alpha_{i,t}\beta_{j,t}\gamma_t+\alpha_{j,t}\beta_{i,t}]$$

$$\times \prod_{t=1}^{380}\frac{\kappa^{\kappa}}{\Gamma(\kappa)}\epsilon_t^{\kappa-1}\exp(-\kappa\epsilon_t)\left[\epsilon_t\alpha_{i,t}\beta_{j,t}\gamma_t\right]^{x_t}\times\left[\epsilon_t\alpha_{j,t}\beta_{i,t}\right]^{y_t}$$

The likelihood marginalised over the random effect is

$$L(\boldsymbol{\alpha},\boldsymbol{\beta},\gamma,\kappa) = \int_{\boldsymbol{\epsilon}} f(\mathbf{x},\mathbf{y},\boldsymbol{\epsilon} \mid \boldsymbol{\alpha},\boldsymbol{\beta},\Gamma,\kappa)p(\boldsymbol{\epsilon}\mid\kappa)d\epsilon$$

$$= \prod_{t=1}^{380}\frac{\Gamma(\kappa+x_t+y_t)}{\Gamma(\kappa)\Gamma(x_t+1)\Gamma(y_t+1)}p_t^{x_t}q_t^{y_t}(1-p_t-q_t)^{\kappa_t}$$

where $p_t=\frac{\mu_t}{\kappa+\mu_t+\lambda_t}$ and $q_t=\frac{\lambda_t}{\kappa+\mu_t+\lambda_t}$, $\mu_t=\alpha_{i,t}\beta_{j,t}\gamma_t$ and $\lambda_t=\alpha_{j,t}\beta_{i,t}$.

Over-dispersion and positive correlations

This is the bivariate Negative Binomial model which is able to explain both over-dispersion and positive correlation. The marginal variance, covariance and correlations of the bivariate distribution can be derived by using standard identities.

$$\mathrm{Var}(X_t) = \mu_t + \frac{\mu_t^2}{\kappa}$$

$$\mathrm{Cov}(X_t, Y_t) = \frac{\mu_t \lambda_t}{\kappa}$$

$$\mathrm{Cor}(X_t, Y_t) = \frac{\mu_t \lambda_t}{\sqrt{(\kappa \mu_t + \mu_t^2)(\kappa \lambda_t + \lambda_t^2)}}$$

where μ_t is the expected score of the home side and λ_t is the expected score of the away side.

Sequential updating of a state space model

Sequential inference

We assume that the dynamic parameters $\boldsymbol{\theta}_t = \{\boldsymbol{\alpha}_t, \boldsymbol{\beta}_t, \gamma_t\}$ evolve over time in such away that there is a loss of information from posterior to prior through the use of discount parameters. Gamerman et al.,(2013) and others express a conjugate transition equation to the conditionally Poisson likelihood as

$$\begin{pmatrix} X_t \\ Y_t \end{pmatrix} \mid \kappa \sim \mathrm{Poisson}\begin{pmatrix} \alpha_{i,t}\beta_{j,t}\gamma_t\epsilon_t \\ \alpha_{j,t}\beta_{i,t}\epsilon_t \end{pmatrix} \qquad \text{(Observation Equation)}$$

$$\frac{\omega\boldsymbol{\theta}_t}{\boldsymbol{\theta}_{t-1}} \sim \mathrm{Beta}\left(\omega p^{\theta}_{t-1}, (1-\omega)p^{\theta}_{t-1}\right) \qquad \text{(State Equation)}$$

We will see that $0 < \omega \leq 1$ has the effect of discounting the posterior from the previous observation.

Sequential Bayesian inference

Posterior | Prior | Posterior

$$\theta_{t-1} \mid X_{1:t-1}, Y_{1:t-1} \longrightarrow \tilde{\theta}_t \mid X_{1:t-1}, Y_{1:t-1} \longrightarrow \theta_t \mid X_{1:t}, Y_{1:t}$$

$$\tilde{\theta}_t = \omega\theta_{t-1}$$

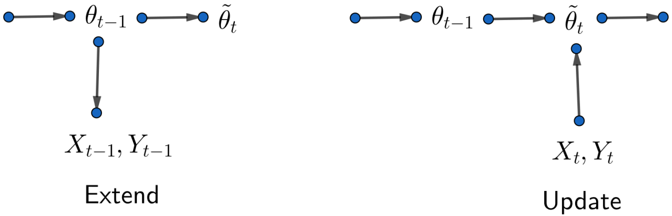

Figure: Sequential Bayesian inference consists of an extend an an update

$\text{Posterior}_{t-1} \xrightarrow{Extend} \text{Prior}_t \xrightarrow{Update} \text{Posterior}_t$

The extend argument below follows from induction.

$$\theta_{t-1} \mid x_{1:t-1}, y_{1:t-1} \sim \text{Gamma}\left(p^{\theta}_{t-1}, q^{\theta}_{t-1}\right) \qquad \text{(Posterior)}$$

$$\pi(\theta_t \mid x_{1:t-1}, y_{1:t-1}) \propto \int_{\theta_{t-1}} \pi(\theta_{t-1} \mid x_{1:t-1}, y_{1:t-1})\, \pi(\theta_t \mid \theta_{t-1}) d\theta_{t-1} \qquad \text{(Extend)}$$

$$\begin{aligned} \theta_t \mid x_{1:t-1}, y_{1:t-1} &\sim \text{Gamma}\left(\omega p^{\theta}_{t-1}, \omega q^{\theta}_{t-1}\right) \qquad \text{(Prior)} \\ &\sim \text{Gamma}\left(\tilde{p}^{\theta}_{t}, \tilde{q}^{\theta}_{t}\right) \end{aligned}$$

$$\pi(\theta_t \mid x_{1:t}, y_{1:t}) \propto \pi(\theta_t \mid x_{1:t-1}, y_{1:t-1}) p(x_t, y_t \mid \theta_t) \qquad \text{(Update)}$$

$$\theta_t \mid x_{1:t}, y_{1:t} \sim \text{Gamma}\left(p^{\theta}_{t}, q^{\theta}_{t}\right) \qquad \text{(Posterior)}$$

and the bi-product of the update step is the predictive distribution:

$$\begin{aligned} p(x_t, y_t \mid x_{1:t-1}, y_{1:t-1}) &= \int_{\theta_t} p(x_t, y_t \mid \theta_t) \pi(\theta_t \mid x_{1:t-1}, y_{1:t-1}) d\theta_t \\ &= \frac{1}{x_t! y_t!} \times \frac{\Gamma(p^{\theta}_t)}{\Gamma(\tilde{p}^{\theta}_t)} \frac{{\tilde{q}^{\theta}_t}^{\tilde{p}^{\theta}_t}}{{q^{\theta}_t}^{p^{\theta}_t}} \end{aligned}$$

The Extend step (before each game)

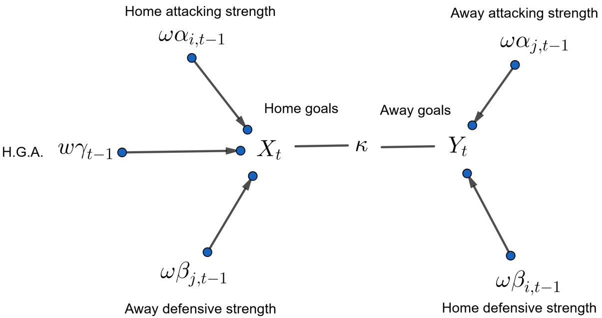

Figure: The extend step takes the posterior from the last observation to create new prior. It expresses loss of information by discounting or down weighting past form The discount factor, ω, preserves the mean of the posterior but adds uncertainty to the priory by increasing its variance.

Step 1 Football model: Extending the previous posterior

The priors in terms of the previous posterior

	Prior	Posterior
Attack	$\alpha_{i,t} \sim \text{Gamma}\left(\tilde{p}^{\alpha}_{i,t}, \tilde{q}^{\alpha}_{i,t}\right)$	$\alpha_{i,t} \sim \text{Gamma}\left(p^{\alpha}_{i,t}, q^{\alpha}_{i,t}\right)$
Defense	$\beta_{i,t} \sim \text{Gamma}\left(\tilde{p}^{\beta}_{i,t}, \tilde{q}^{\beta}_{i,t}\right)$	$\beta_{i,t} \sim \text{Gamma}\left(p^{\beta}_{i,t}, q^{\beta}_{i,t}\right)$
HGA	$\gamma_t \sim \text{Gamma}\left(\tilde{p}^{\gamma}_{t}, \tilde{q}^{\gamma}_{t}\right)$	$\gamma_t \sim \text{Gamma}\left(p^{\gamma}_{t}, q^{\gamma}_{t}\right)$

$$
\begin{aligned}
\alpha_{i,t} &\sim \text{Gamma}\left(\tilde{p}^{\alpha}_{i,t}, \tilde{q}^{\alpha}_{i,t}\right), & \tilde{p}^{\alpha}_{i,t} &= \omega p^{\alpha}_{i,t-1}, & \tilde{q}^{\alpha}_{i,t} &= \omega q^{\alpha}_{i,t-1} & &(\text{AS}\quad\text{Home})\\
\alpha_{j,t} &\sim \text{Gamma}\left(\tilde{p}^{\alpha}_{j,t}, \tilde{q}^{\alpha}_{j,t}\right), & \tilde{p}^{\alpha}_{j,t} &= \omega p^{\alpha}_{j,t-1}, & \tilde{q}^{\alpha}_{j,t} &= \omega q^{\alpha}_{j,t-1}, & &(\text{AS Away})\\
\beta_{j,t} &\sim \text{Gamma}\left(\tilde{p}^{\beta}_{j,t}, \tilde{q}^{\beta}_{j,t}\right), & \tilde{p}^{\beta}_{j,t} &= \omega p^{\beta}_{j,t-1}, & \tilde{q}^{\beta}_{j,t} &= \omega q^{\beta}_{j,t-1}, & &(\text{DS Away})\\
\beta_{i,t} &\sim \text{Gamma}\left(\tilde{p}^{\beta}_{i,t}, \tilde{q}^{\beta}_{i,t}\right), & \tilde{p}^{\beta}_{i,t} &= \omega p^{\beta}_{i,t-1}, & \tilde{q}^{\beta}_{i,t} &= \omega q^{\beta}_{i,t-1}, & &(\text{DS}\quad\text{Home})\\
\gamma_t &\sim \text{Gamma}\left(\tilde{p}^{\gamma}_{t}, \tilde{q}^{\gamma}_{t}\right), & \tilde{p}^{\gamma}_{t} &= w_t p^{\gamma}_{t-1}, & \tilde{q}^{\gamma}_{t} &= w_t q^{\gamma}_{t-1}, & &(\text{HGA})\\
\epsilon_t &\sim \text{Gamma}\left(\kappa, \kappa\right), & & & & & &(\text{R.E})
\end{aligned}
$$

The mean is preserved but the variance has increased.

The Update step (after each game)

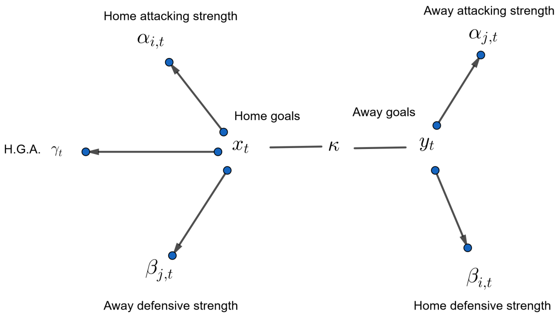

Figure: The Update step is just an application of Bayes theorem. The prior is updated to encompass information from the score of the latest football game. The normalising constant of Bayes theorem yields the predictive distribution which give us the cumulative evidence, needed for model comparisons.

The full conditional posteriors (after each game)

We denote the fixed parameters by ϕ. The dynamic posterior parameters $\boldsymbol{\theta}_t = \{\alpha_{j,t}, \alpha_{i,t}, \beta_{i,t}, \beta_{j,t}\gamma_t, \epsilon_t\ \}$ are updated at the end of a game at time t using the following joint posterior

$$
\begin{aligned}
\pi(\boldsymbol{\theta}_t \mid \mathbf{x}_{1:t}, \mathbf{y}_{1:t}, \phi) \propto\ & \underbrace{\exp -[\epsilon_t(\alpha_{i,t}\beta_{j,t}\gamma + \alpha_{j,t}\beta_{i,t})]\ \times [\epsilon_t\alpha_{i,t}\beta_{j,t}\gamma]^{x_t} \times [\epsilon_t\alpha_{j,t}\beta_{i,t}]^{y_t}}_{\text{Likelihood}} \\
& \times \underbrace{\alpha_{i,t}^{\tilde{p}^{\alpha}_{i,t}-1} \exp(-\tilde{p}^{\alpha}_{i,t}\alpha_{i,t})\ \alpha_{j,t}^{\tilde{p}^{\alpha}_{j,t}-1} \exp(-\tilde{p}^{\alpha}_{j,t}\alpha_{j,t})}_{\text{Prior Attacking Strengths}} \\
& \times \underbrace{\beta_{i,t}^{\tilde{p}^{\beta}_{i,t}-1} \exp(-\tilde{p}^{\beta}_{i,t}\beta_{i,t})\ \beta_{j,t}^{\tilde{p}^{\beta}_{j,t}-1} \exp(-\tilde{p}^{\beta}_{j,t}\beta_{j,t})}_{\text{Prior Defensive Strengths}} \\
& \times \underbrace{\gamma_t^{\tilde{p}^{\gamma}_t\, t-1} \exp(-\tilde{q}^{\gamma}_t\gamma_t)}_{\text{HGA}} \times \underbrace{\epsilon_t^{\kappa-1} \exp(-\kappa\epsilon_t)}_{\text{Random effect}}
\end{aligned}
$$

Step 2 Updating the sufficient statistics

The updates can be formulated by examining the full conditional posterior distributions. All the dynamic parameters have known shape parameters. However the scale parameters involve other parameters and are not known. However the expectations at the previous iteration can be used as a proxy for a sufficient statistic.

The update equations are then

$$
\begin{aligned}
p^{\alpha}_{i,t} &\leftarrow \tilde{p}^{\alpha}_{i,t} + x_t & q^{\alpha}_{i,t} &\leftarrow \tilde{q}^{\alpha}_{i,t} + \hat{\gamma}_t\hat{\beta}_{j,t}, && \text{(AS home)}\\
p^{\alpha}_{j,t} &\leftarrow \tilde{p}^{\alpha}_{j,t} + y_t & q^{\alpha}_{j,t} &\leftarrow \tilde{q}^{\alpha}_{j,t} + \hat{\beta}_{i,t}, && \text{(AS away)}\\
p^{\beta}_{i,t} &\leftarrow \tilde{p}^{\beta}_{i,t} + y_t & q^{\beta}_{i,t} &\leftarrow \tilde{q}^{\beta}_{i,t} + \hat{\alpha}_{j,t}, && \text{(DS home)}\\
p^{\beta}_{j,t} &\leftarrow \tilde{p}^{\beta}_{j,t} + x_t & q^{\beta}_{j,t} &\leftarrow \tilde{q}^{\beta}_{j,t} + \hat{\gamma}_t\hat{\alpha}_{i,t} && \text{(DS away)}\\
p^{\gamma}_{t} &\leftarrow \tilde{p}^{\gamma}_{t} + x_t & q^{\gamma}_{t} &\leftarrow \tilde{q}^{\gamma}_{t} + \hat{\alpha}_{i,t}\hat{\beta}_{j,t} && \text{(HGA)}\\
p^{\epsilon}_{t} &\leftarrow \kappa + x_t + y_t & q^{\epsilon}_{t} &\leftarrow \kappa + \hat{\gamma}\hat{\alpha}_{i,t}\hat{\beta}_{j,t} + \hat{\alpha}_{j,t}\hat{\beta}_{i,t} && \text{(RE)}
\end{aligned}
$$

where $\hat{\alpha}_{i,t} = \frac{p^{\alpha}_{i,t}}{q^{\alpha}_{i,t}}$, $\hat{\beta}_{i,t} = \frac{p^{\beta}_{i,t}}{q^{\beta}_{i,t}}$ and $\hat{\gamma}_t = \frac{p^{\gamma}_t}{q^{\gamma}_t}$.

Updating the cumulative evidence (after each game)

The predictive can be calculated as

$$p(x_t, y_t \mid x_{1:t-1}, y_{1:t-1}, \phi) = \frac{1}{x_t! y_t!} \times \frac{\Gamma(p^{\alpha}_{i,t})}{\Gamma(\tilde{p}^{\alpha}_{i,t})} \frac{\tilde{q}^{\alpha\ \tilde{p}^{\alpha}_{i,t}}_{i,t}}{q^{\alpha\ p^{\alpha}_{i,t}}_{i,t}} \times \frac{\Gamma(p^{\alpha}_{j,t})}{\Gamma(\tilde{p}^{\alpha}_{i,t})} \frac{\tilde{q}^{\alpha\ \tilde{p}^{\alpha}_{j,t}}_{j,t}}{q^{\alpha\ p^{\alpha}_{j,t}}_{j,t}}$$
$$\times \frac{\Gamma(p^{\beta}_{i,t})}{\Gamma(\tilde{p}^{\beta}_{i,t})} \frac{\tilde{q}^{\beta\ \tilde{p}^{\beta}_{i,t}}_{i,t}}{q^{\beta\ p^{\beta}_{i,t}}_{i,t}}' \times \frac{\Gamma(p^{\beta}_{j,t})}{\Gamma(\tilde{p}^{\beta}_{i,t})} \frac{\tilde{q}^{\beta\ \tilde{p}^{\beta}_{j,t}}_{j,t}}{q^{\beta\ p^{\beta}_{j,t}}_{j,t}} \times \frac{\Gamma(p^{\gamma}_{t})}{\Gamma(\tilde{p}^{\gamma}_{t})} \frac{\tilde{q}^{\gamma\ \tilde{p}^{\gamma}_{t}}_{t}}{q^{\gamma\ p^{\gamma}_{t}}_{t}} \times \frac{\Gamma(p^{\epsilon}_{t})}{\Gamma(\kappa)} \frac{\kappa^{\kappa}}{q^{\epsilon\ p^{\epsilon}_{t}}_{t}}. \quad (1)$$

The evidence for can be calculated as

$$Z_{k,t} = \prod_{t^*=1}^{t} p(x_{t^*}, y_{t^*} \mid x_{1:t^*-1}, y_{1:t^*-1}, \phi_k) \quad (2)$$

Here $\phi_k, \quad k = 1, 2, \ldots, K$ denotes a particular configuration of the fixed parameters which include ω and κ, the discount and correlation parameters.

Dynamic model averaging

Dynamic model averaging

We now look at the online evolution of the mixture of models.
We set a uniform prior on each configuration of fixed parameters ϕ_k at the beginning of the season as $\pi(\phi_k) = \frac{1}{K}$. The posterior weights of each component of the mixture, $\Omega_{k,t}$, are given by

$$\Omega_{k,t} = \pi(\phi_{k,t} \mid x_{1:t}, y_{1:t}) = \frac{p(x_t, y_t \mid x_{1:t-1}, y_{1:t-1}, \phi_k)}{\sum_{k=1}^{K} p(x_t, y_t \mid x_{1:t-1}, y_{1:t-1}, \phi_k)}.$$

The filtered posteriors of the attacking and defensive strengths are now a mixture of Gamma distributions

$$\boldsymbol{\alpha}_t \mid \phi = \sum_{k=1}^{K} \Omega_{k,t} \boldsymbol{\alpha}_{k,t}$$

where $\boldsymbol{\alpha}_{k,t} \sim \mathrm{Gamma}\left(\mathbf{p}^{\alpha}_{k,t}, \mathbf{q}^{\alpha}_{k,t}\right)$, for $k = 1, 2, \ldots, K$

Sequential prediction from the dynamic mixture

The cumulative evidence for the mixture of models is the sum of the predictive distributions weighetd by the cumulative evidence up to last observation.

$$Z_t^* = \prod_{t^*=1}^{t} \sum_{k=1}^{K} \Omega_{k,t-1} p(x_{t^*}, y_{t^*} \mid x_{1:t^*-1}, y_{1:t^*-1}, \phi_k) \qquad (3)$$

We show that the cumulative predictive performance of the mixture is usually much better than the cumulative predictive performance any of the components of the mixture.

The plots on the next few slides contrast the log evidence of the mixture, Z_t^* to the log evidence of each of the components of the mixture $Z_{k,t}$ $k = 1, 2, \ldots 20$.

Weighted prediction of a finite number of predictive distributions

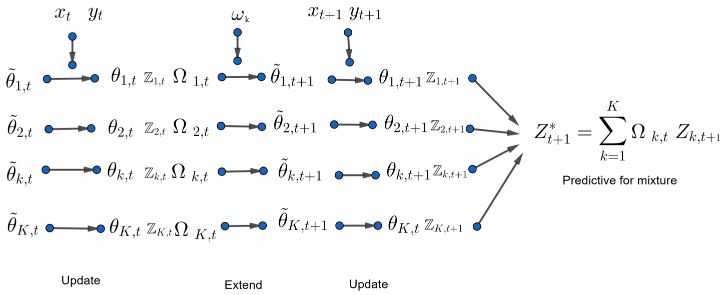

Figure: In model averaging the predictive from K models is combined and weighted by the cumulative evidence from previous step

Comparison of evidence for models for 1999 Season

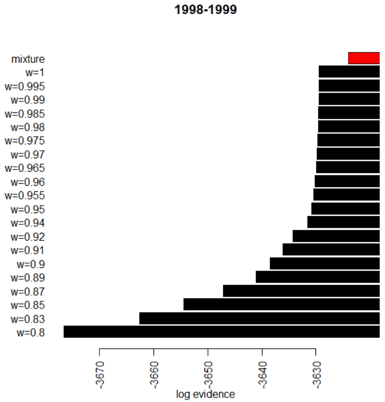

Figure: The log evidence of the mixture compared to the log evidence of each of the components of the mixture

Comparison of evidence for models for 2001 Season

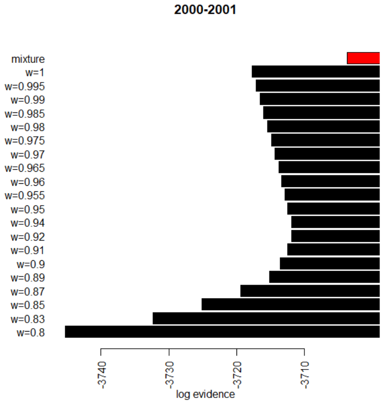

Figure: The log evidence of the mixture compared to the log evidence of each of the components of the mixture

Comparison of evidence for models for 2004 Season

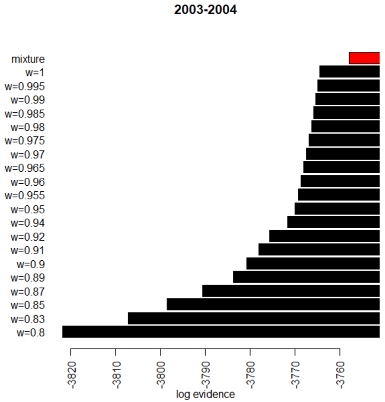

Figure: The log evidence of the mixture compared to the log evidence of each of the components of the mixture

Comparison of evidence for models for 2005 Season

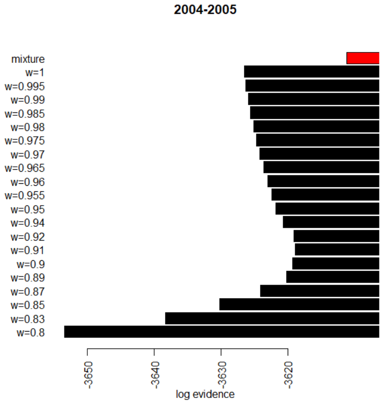

Figure: The log evidence of the mixture compared to the log evidence of each of the components of the mixture

Comparison of evidence for models for 2009 Season

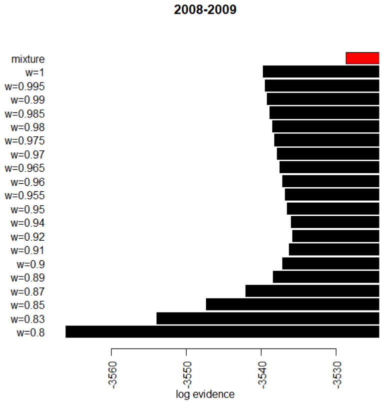

Figure: The log evidence of the mixture compared to the log evidence of each of the components of the mixture

Comparison of evidence for models for 2015 Season

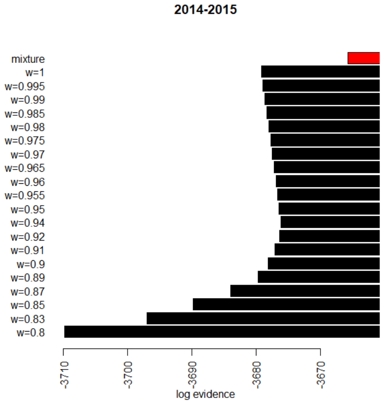

Figure: The pedictivity of the mixture compared to the log evidence of each of the components of the mixture

Comparison of evidence for models for 2019 Season

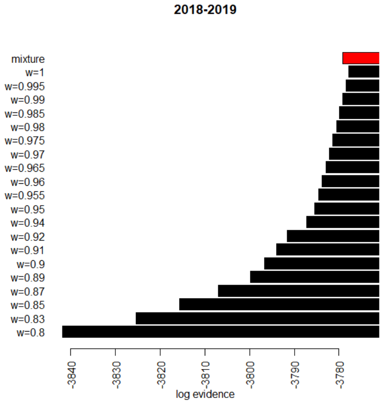

Figure: The log evidence of the mixture compared to the log evidence of each of the components of the mixture

The evolution of posterior weights with different discount factors

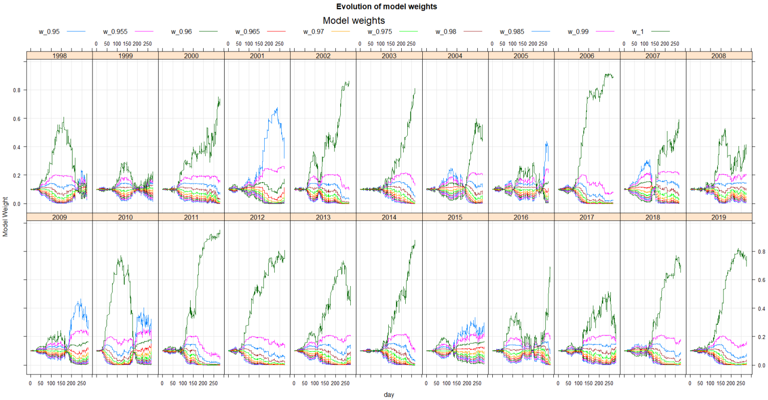

Figure: Sequential posterior weights within and between seasons. $\omega \in \{0.95, 0.995, 0.96, 0.965, 0.97, 0.975, 0.98, 0.985, 0.99, 1\}$

The evolution of weights of models with different values of κ and ω.

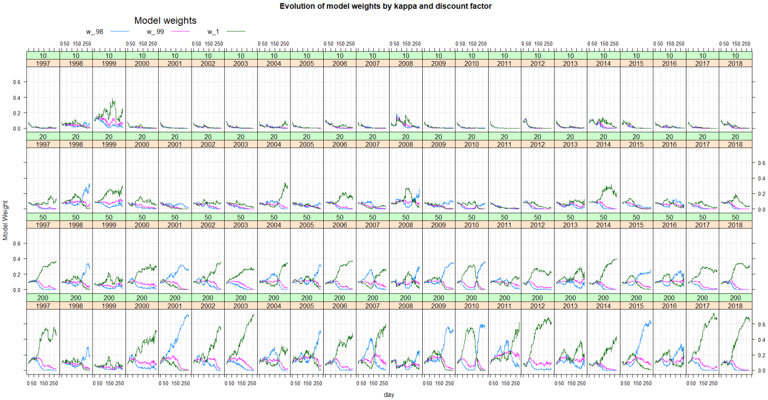

Figure: Sequential posterior weights within and between seasons. $\kappa \in \{10, 20, 50, 200\}$ and $\omega \in \{0.98, 0.99, 1\}$

Filtered mixtures of abilities for Chelsea and Man United

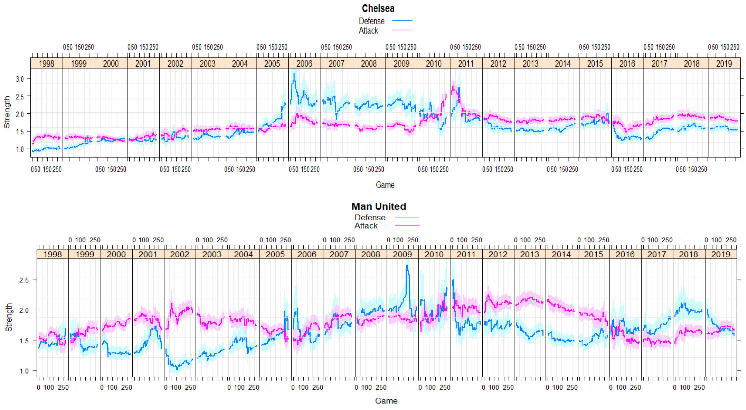

Figure: The evolution of form of Chelsea (Winners in 05,06,10,15,17) and Man United (Winners in 99,00,01,03,07,08,09,11,13)

Filtered mixtures of abilities for Arsenal and Liverpool

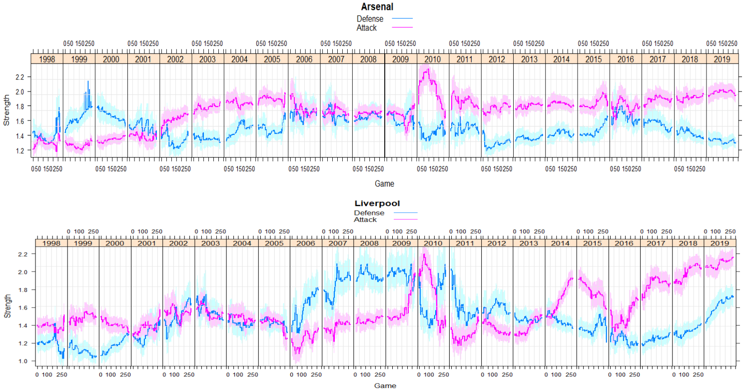

Figure: The evolution of form of Arsenal (98,02,04) and Liverpool.

BK

Conclusions

- MCMC and simulation based methodologies are of limited use for fast online prediction.
- Online sequential updating and model averaging give superior predictions
- Constant exponential smoothing with a fixed rate parameter is not appropriate for football modelling
- A small varying amount of over dispersion and positive correlation can be dealt with by model averaging.
- Mixtures of models highlight sharp changes in form and result in better predictions.
- Embrace conjugacy when you can

Winners

	Season	Winners
1	1997/98	Arsenal
2	1998/99	Manchester United
3	1999/00	Manchester United
4	2000/01	Manchester United
5	2001/02	Arsenal
6	2002/03	Manchester United
7	2003/04	Arsenal
8	2004/05	Chelsea
9	2005/06	Chelsea
10	2006/07	Manchester United
11	2007/08	Manchester United
12	2008/09	Manchester United
13	2009/10	Chelsea
14	2010/11	Manchester United
15	2011/12	Manchester City
16	2012/13	Manchester United
17	2013/14	Manchester City
18	2014/15	Chelsea
19	2015/16	Leicester City
20	2016/17	Chelsea
21	2017/18	Manchester City
22	2018/19	Manchester City