Forecasting match results in European football competitions: new dynamic models and comparisons

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

(a) VU Amsterdam and Tinbergen Institute
(b) CREATES, Aarhus University

CMStatistics 2016

10 December 2016

1. Modelling football matches: which approach?
2. Static, semi-dynamic, or dynamic models
3. Methodology of new dynamic model
4. Forecasting study
5. Conclusions

Main contributions of football models divided into nine categories.

G (Goals): match results are treated as pairwise observations.
D (Difference): margin of victory is modelled (difference between home and away goals).
T (Toto): win, draw, and loss probabilities are modelled directly.

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

Main contributions of this paper

1. Determine which distribution performs best in forecasting.

Main contributions of this paper

1. Determine which distribution performs best in forecasting.
2. Determine if dynamic models are preferred above others.

Main contributions of this paper

1. Determine which distribution performs best in forecasting.
2. Determine if dynamic models are preferred above others.
3. Develop a dynamic model that, despite the high dimensionality, is computationally fast and performs well in forecasting.

Main contributions of this paper

1. Determine which distribution performs best in forecasting.
2. Determine if dynamic models are preferred above others.
3. Develop a dynamic model that, despite the high dimensionality, is computationally fast and performs well in forecasting.
4. Perform an extensive forecasting study where several competitions for many years are forecasted.

The models in the category 'Goals' typically use a Bivariate Poisson distribution for the pair of counts (X, Y)

$$\mathrm{P}(X=x,\, Y=y) = \exp\left(-\lambda_1-\lambda_2-\lambda_3\right) \frac{\lambda_1^x}{x!}\, \frac{\lambda_2^y}{y!} \sum_{k=0}^{min(x,y)} \binom{x}{k}\binom{y}{k} k! \left(\frac{\lambda_3}{\lambda_1\, \lambda_2}\right)^k ,$$

where X is the number of home goals and Y the number of away goals. We have intensities $\lambda_1, \lambda_2 > 0$ and covariance between counts $\lambda_3 \geq 0$.

It can be shown that

$$\begin{aligned} \mathrm{E}(X) = \mathrm{Var}(X) &= \lambda_1 + \lambda_3, \\ \mathrm{E}(Y) = \mathrm{Var}(Y) &= \lambda_2 + \lambda_3, \\ \mathrm{Cov}(X,\, Y) &= \lambda_3. \end{aligned}$$

For $\lambda_3 = 0$, we have (X, Y) distributed by a double Poisson.

We are often interested in the probabilities of a home win, draw, or away win which are in the 'pairwise observation case' given by

1. $\mathrm{P}(\text{home}) = \mathrm{P}(X > Y)$
2. $\mathrm{P}(\text{draw}) = \mathrm{P}(X = Y)$
3. $\mathrm{P}(\text{away}) = \mathrm{P}(X < Y)$

The intensities λ_1, λ_2 are expressed as functions of a latent strenght of attack (α) and defence (β)

$$\lambda_1 = \exp(\delta + \alpha_1 - \beta_2),$$
$$\lambda_2 = \exp(\alpha_2 - \beta_1),$$

see Maher(1982) and where δ is the home ground advantage.

The models in the category ‘Difference’ typically use the Skellam distribution for the difference between the pair of counts (X, Y).

For a pair of counts (X, Y) that is distributed by a bivariate Poisson, possibly with $\lambda_3 \neq 0$, we have that

$$Z = X - Y \sim \text{Skellam}(\lambda_1, \lambda_2),$$

with

$$\mathrm{P}(Z = z) = \exp\left(-\lambda_1 - \lambda_2\right) \left(\frac{\lambda_1}{\lambda_2}\right)^{z/2} I_{|z|}(2\sqrt{\lambda_1 \lambda_2}),$$

for support $z \in \mathbb{Z} = \{0, \pm 1, \pm 2, \ldots\}$ and where $I_\nu(z)$ is the modified Bessel function of order ν, see Skellam (1946).

We have

$$\mathrm{E}(Z) = \lambda_1 - \lambda_2, \qquad \mathrm{Var}(Z) = \lambda_1 + \lambda_2.$$

The probabilities of a home win, draw, or away win are in the 'margin of victory case' given by

1. $\mathrm{P}(\text{home}) = \mathrm{P}(Z > 0)$
2. $\mathrm{P}(\text{draw}) = \mathrm{P}(Z = 0)$
3. $\mathrm{P}(\text{away}) = \mathrm{P}(Z < 0)$

Also for this category, the intensities λ_1, λ_2 are expressed as functions of a latent strenght of attack (α) and defence (β)

$$\lambda_1 = \exp(\delta + \alpha_1 - \beta_2),$$
$$\lambda_2 = \exp(\alpha_2 - \beta_1).$$

The models in the category ‘Toto’ typically use an ordered probit or ordered logit model.

The categories are home win (0), draw (1), and away win (2).

In the ordered probit model we have

$$\begin{aligned}
\mathrm{P}(y_{i,j} = 0) &= \Phi(\kappa_1 + \theta_i - \theta_j), \\
\mathrm{P}(y_{i,j} = 1) &= \Phi(\kappa_2 + \theta_i - \theta_j) - \Phi(\kappa_1 + \theta_i - \theta_j), \\
\mathrm{P}(y_{i,j} = 2) &= 1 - \Phi(\kappa_2 + \theta_i - \theta_j),
\end{aligned}$$

for a match between home team i and away team j where κ_1, κ_2 are the cutoff points and θ_i is the total strength of team i.

κ_1 corresponds with home ground advantage.

Evolvement of parameters over time

Main contributions of football models divided into nine categories.

G (Goals): match results are treated as pairwise observations.
D (Difference): margin of victory is modelled (difference between home and away goals).
T (Toto): win, draw, and loss probabilities are modelled directly.

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

Static: model parameters are static over time. Less realistic but very easy and fast in estimating and forecasting.

Static: model parameters are static over time. Less realistic but very easy and fast in estimating and forecasting.

Semi-dynamic: the influence of match results from the more distant past is downgraded. In a likelihood setting: the likelihood is raised to the power of a time function and certain match results are downgraded more than others.

$$L(\theta; y) = \prod_{i=1}^{n} \{f(y_i; \theta)\}^{\phi(t)}$$

Ad-hoc but more realistic than static parameters. Fast estimation and forecasting however tuning of the ‘time-function’ can be problematic.

Static: model parameters are static over time. Less realistic but very easy and fast in estimating and forecasting.

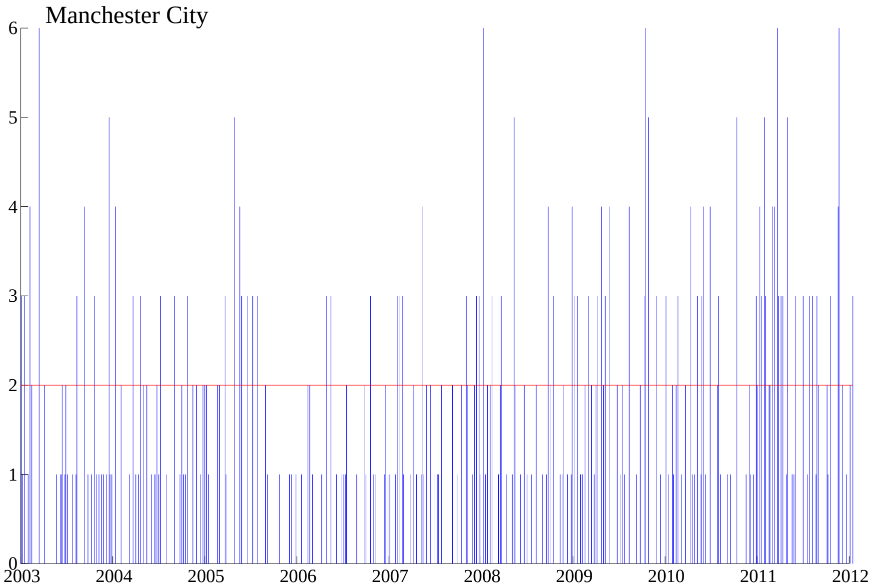
Semi-dynamic: the influence of match results from the more distant past is downgraded. In a likelihood setting: the likelihood is raised to the power of a time function and certain match results are downgraded more than others.

$$L(\theta; y) = \prod_{i=1}^{n} \{f(y_i; \theta)\}^{\phi(t)}$$

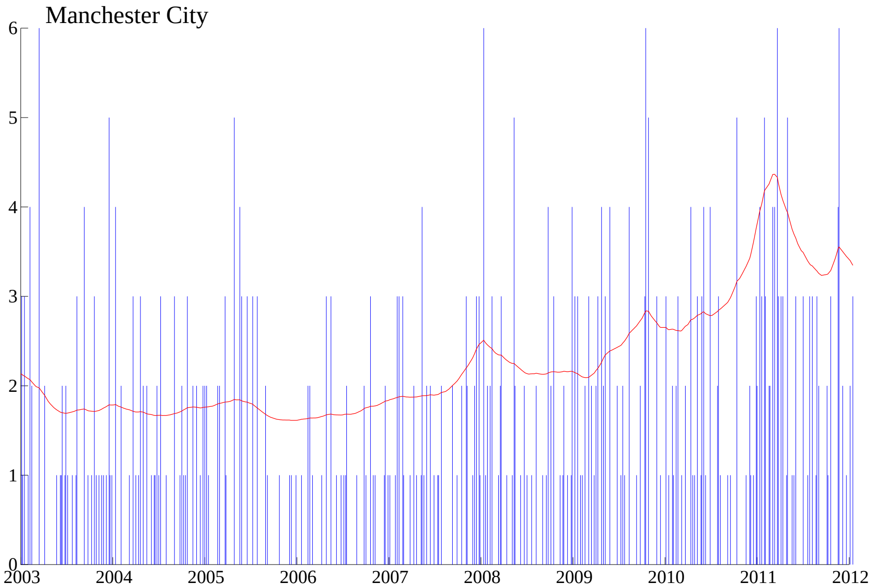
Ad-hoc but more realistic than static parameters. Fast estimation and forecasting however tuning of the 'time-function' can be problematic.

Dynamic: realistic behaviour of model parameters over time. Due to high dimensionality (number of teams) parsimony becomes important. If parameters evolve stochastically over time, estimating the model can be involved due to the high dimensional signal.

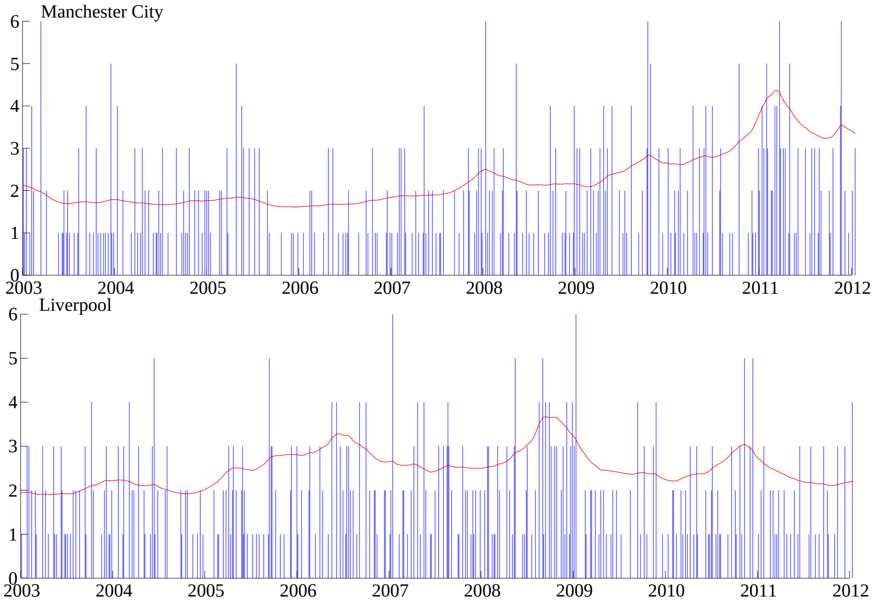
Static strength of attack, Manchester City, 2003-2012



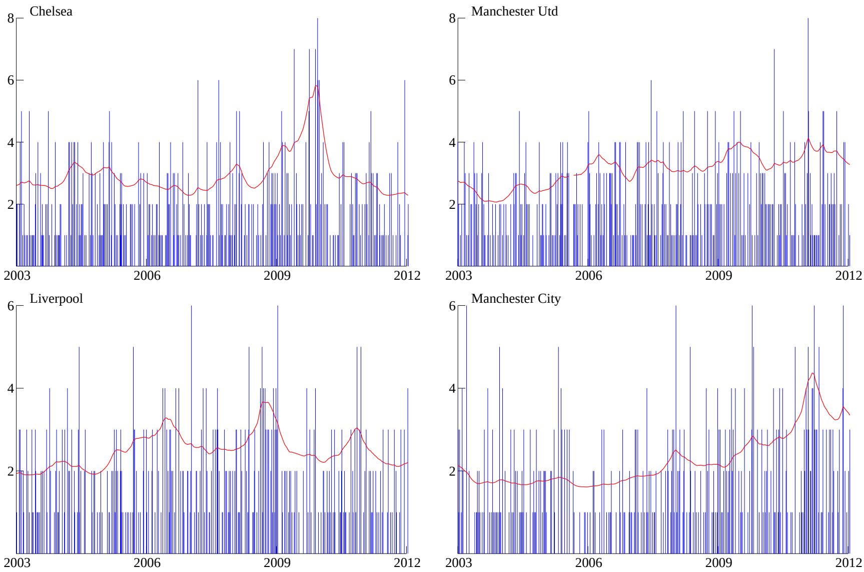
Dynamic strength of attack, Manchester City, 2003-2012



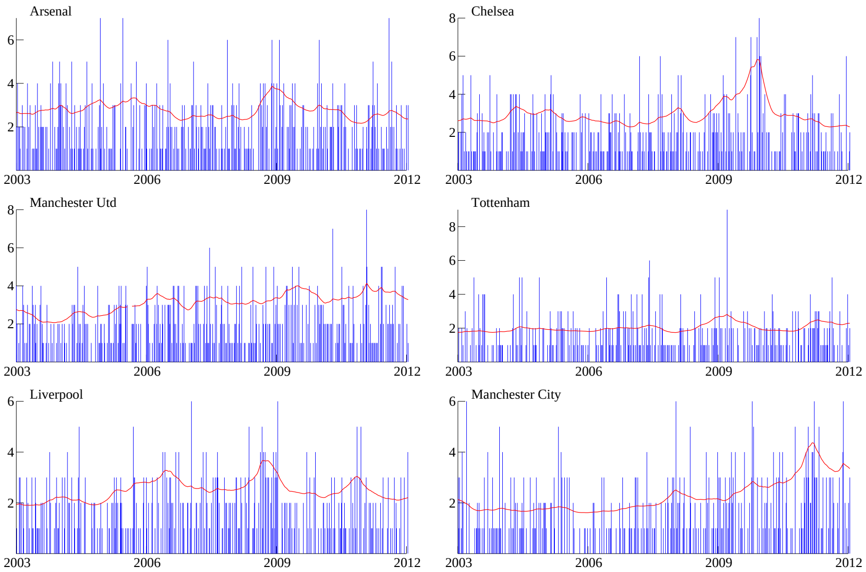
Strength of attack Manchester City and Liverpool, 2003-2012



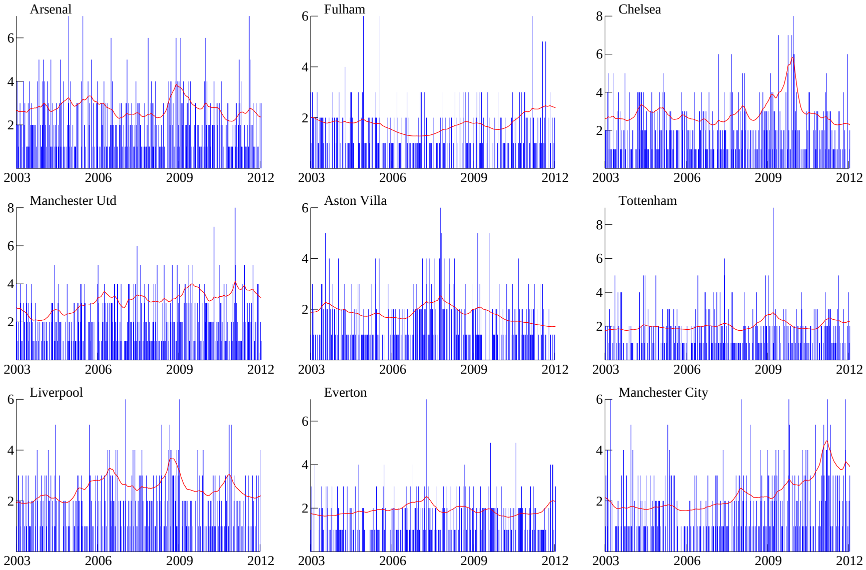
Dynamic strength of attack 4 teams, 2003-2012



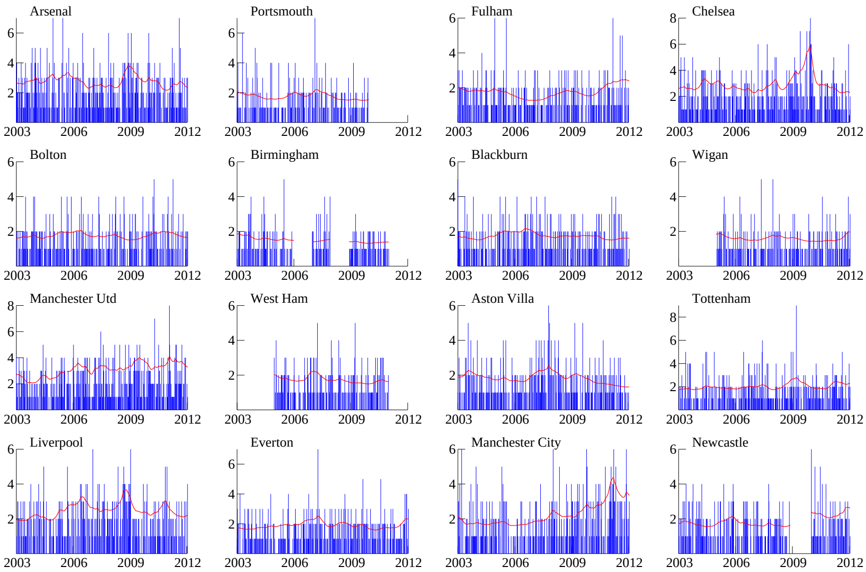
Dynamic strength of attack 6 teams, 2003-2012



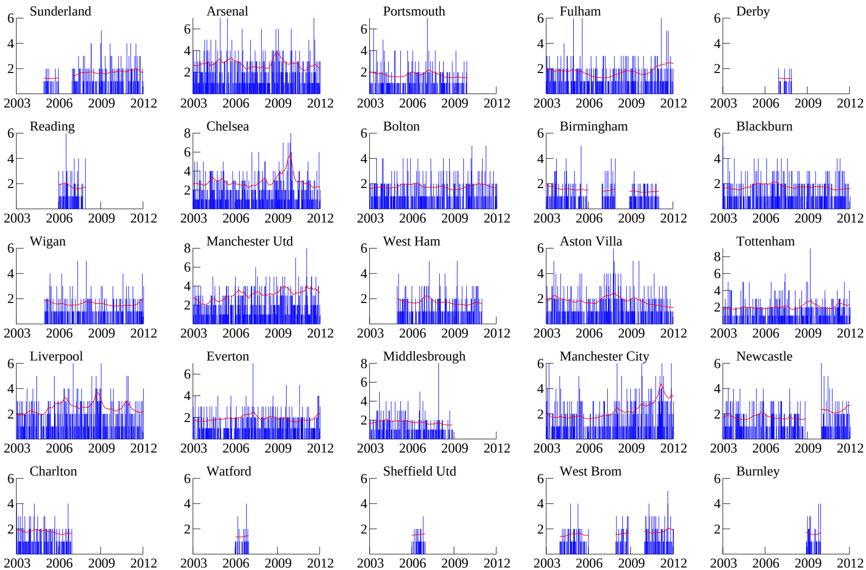
Dynamic strength of attack 9 teams, 2003-2012



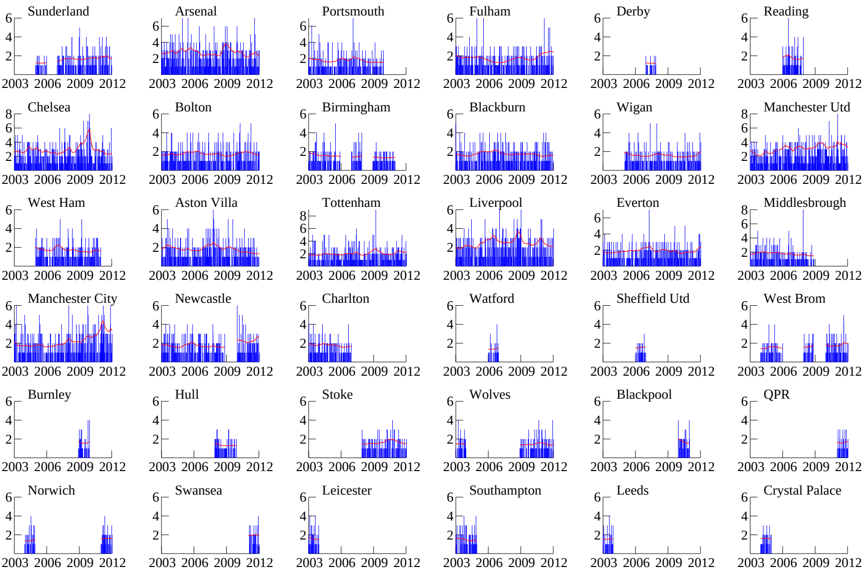
Dynamic strength of attack 16 teams, 2003-2012



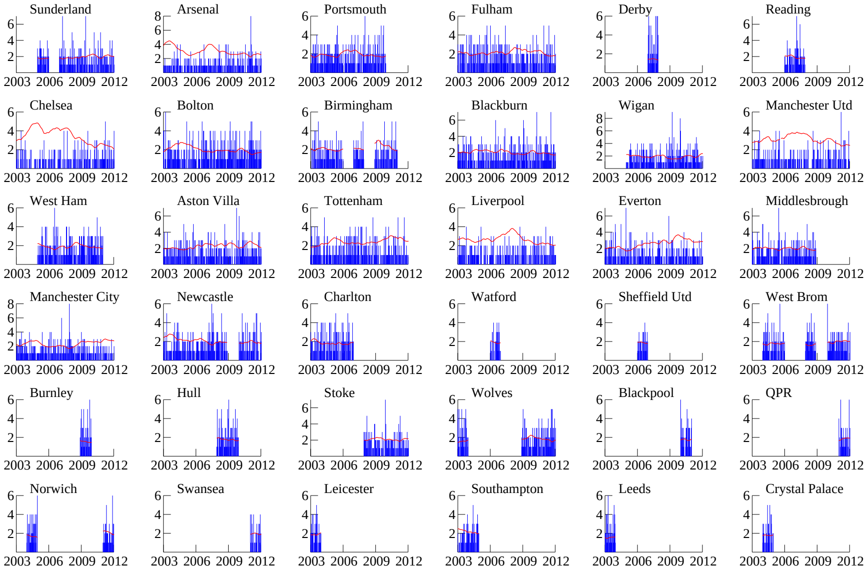
Dynamic strength of attack 25 teams, 2003-2012



Dynamic strength of attack 36 teams, 2003-2012



Dynamic strength of defence 36 teams, 2003-2012



Intensities become dynamic in a high dimensional panel of football teams and match results.

$$\lambda_1 = \lambda_{ijt} = \exp(\delta + \alpha_{it} - \beta_{jt}),$$
$$\lambda_2 = \lambda_{jit} = \exp(\alpha_{jt} - \beta_{it}).$$

1. A team's capability is likely to be best summarized in separate measures of their ability to attack and their ability to defend.
2. The intensity to score goals should be based on the abilities of both teams.
3. The model should take into account the so called 'home ground advantage'.
4. A team's capability is likely to be closely related to it's recent performance.

We allow α_{it} and β_{it} to vary over time by applying the score-driven framework of Creal (2013).

We have N teams and we let the $2N \times 1$ factor f_t consist of all α_{it} and β_{it} so that

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

Furthermore, let y_t be a $N \times 1$ vector of observations which we assume is generated by the observation density

$$y_t \sim p(y_t | f_t, \mathcal{F}_t, \psi),$$

where $\mathcal{F}_t$ is the information set available at time t which consists of lagged variables of observations and time-varying factors and where ψ is a parameter vector with static parameters.

The score-driven update of f_t is given by

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

where ω is a vector of autoregressive constants, the coefficient matrices A and B are possibly dependent on a static parameter vector ψ and s_t is the scaled score vector defined by

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

with $S(\cdot)$ a matrix function to scale the score vector.

A score-driven model updates the factor f_{t+1} in the direction of the steepest increase of the log-density at time t given the current parameter f_t and the data history $\mathcal{F}_t$.

Common choices for S_t are unit scaling, the inverse of the Fisher information matrix, or the square root of the Fisher inverse information matrix.

The latter has the advantage of giving s_t a unit variance since the Fisher information matrix is the variance matrix of the score vector.

The score-driven model has three main advantages:

1. The estimates of the time-varying parameter are optimal in a Kullback-Leibler sense, see Blasques et al. (2014)
2. Since the score-driven models are observation driven, their likelihood is known in closed-form
3. The forecasting performance of these models is comparable to their parameter-driven counterparts, see Koopman et al. (2015).

The matrices A and B of the score updating equation

$$f_{t+1} = \omega + As_t + Bf_t,$$

are diagonal for parsimony so there are no spillover effects.

We have $A = \mathrm{diag}(A_{att,1}, \ldots, A_{att,N}, A_{def,1}, \ldots, A_{def,N})$.

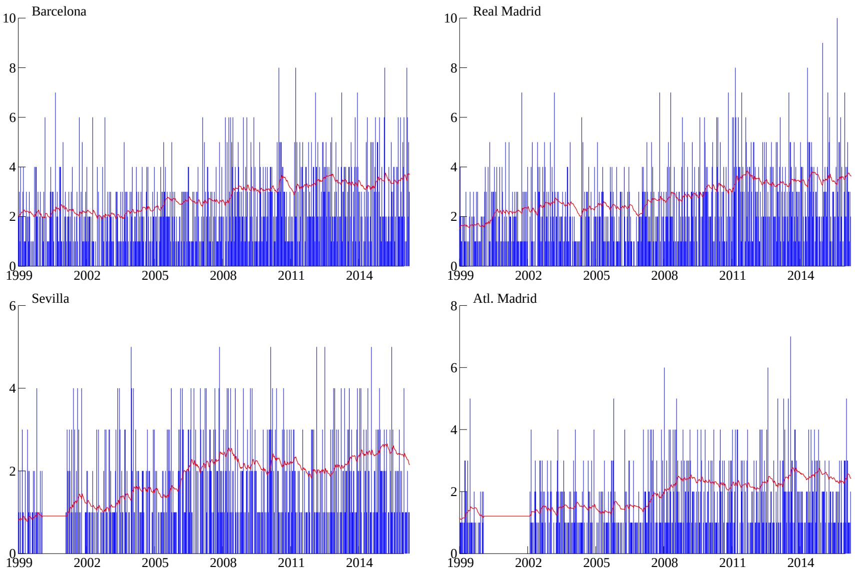
The diagonals of A and B are estimated by maximum likelihood.

Further parsimony is obtained by using a static model and one year of data to obtain f_0 and deriving $\omega = f_0(1 - B)$ from it.

The static parameter vector (estimated by maximum likelihood) is given by

$$\psi = (A_{att}, A_{def}, B_{att}, B_{def}, \delta, \lambda_3).$$

Strength of attack Barcelona, Real Madrid, Sevilla, and Atl. Madrid 1999-2016



1. In a large scale forecasting study we consider six football competitions in Europe for a forecasting period of seven years.

2. We make one-step-ahead forecast with the score driven framework. These are ‘free’ without extra computation since the updating equation is $f_{t+1} = \omega + As_t + Bf_t$.

3. We use an in-sample dataset of ten seasons and let the window expand to use all available data.

4. We compare forecasts with the Diebold Mariano test statistic based on rank probability scores.

Losses based on the average rank probability score (ARPS) obtained by the forecasting of seven seasons of match results for each competition. The lowest ARPS for each column is blue shaded. The benchmark model for the Diebold Mariano (DM) test is the best performing model per competition.

Distribution	Model Specifications	England ARPS	England DM	Germany ARPS	Germany DM	Spain ARPS	Spain DM
Biv. Poisson	Static	0.2060	4.63	0.2149	4.54	0.1951	3.60
Skellam	Static	0.2065	4.90	0.2148	4.49	0.1956	3.78
Ordered probit	Static	0.2084	5.33	0.2158	4.98	0.1959	3.80
Biv. Poisson	Semi-dynamic	0.1999	1.41	0.2111	2.93	0.1912	1.26
Skellam	Semi-dynamic	0.2024	3.27	0.2136	4.09	0.1913	1.36
Ordered probit	Semi-dynamic	0.2009	2.37	0.2123	3.85	0.1922	2.16
Biv. Poisson	Dynamic	0.1994	1.50	0.2083	–	0.1902	–
Skellam	Dynamic	0.1987	–	0.2091	1.36	0.1914	2.39
Ordered probit	Dynamic	0.2008	2.48	0.2099	2.67	0.1914	1.72
Biv. Poisson	Dynamic (SSM)	0.1994	1.52	0.2093	1.44	0.1906	1.58

Losses based on the average rank probability score (ARPS) obtained by the forecasting of seven seasons of match results for each competition. The lowest ARPS for each column is blue shaded. The benchmark model for the Diebold Mariano (DM) test is the best performing model per competition.

Distribution	Model Specifications	Italy ARPS	Italy DM	France ARPS	France DM	Netherlands ARPS	Netherlands DM
Biv. Poisson	Static	0.2039	2.65	0.2104	3.07	0.1974	2.33
Skellam	Static	0.2039	2.56	0.2106	3.18	0.1971	2.16
Ordered probit	Static	0.2046	2.83	0.2110	3.15	0.1969	1.94
Biv. Poisson	Semi-dynamic	0.2023	2.15	0.2072	0.57	0.1938	–
Skellam	Semi-dynamic	0.2025	2.28	0.2075	0.90	0.1940	0.49
Ordered probit	Semi-dynamic	0.2026	2.23	0.2074	0.57	0.1941	0.52
Biv. Poisson	Dynamic	0.2000	–	0.2070	–	0.1945	1.21
Skellam	Dynamic	0.2001	0.25	0.2083	2.03	0.1952	1.92
Ordered probit	Dynamic	0.2011	1.46	0.2071	0.26	0.1944	0.61
Biv. Poisson	Dynamic (SSM)	0.2010	1.53	0.2078	2.10	0.1947	1.15

We developed a new dynamic model for the analysis and forecasting of a high dimensional panel of time series. An extensive forecasting study into the match results of football teams in European competitions allowed us to draw a number of conclusions:

1. The merging of data to obtain simpler models results in worse forecasts. The bivariate Poisson distribution performs for 5 out 6 competitions the best in forecasting so we probably lose information if we merge data.
2. Dynamic models perform better than the static and semi-dynamic counterparts.
3. Score driven models are a good (and much faster) alternative for parameter driven counterparts (state space models, Bayesian analysis).

Thank you for your attention!