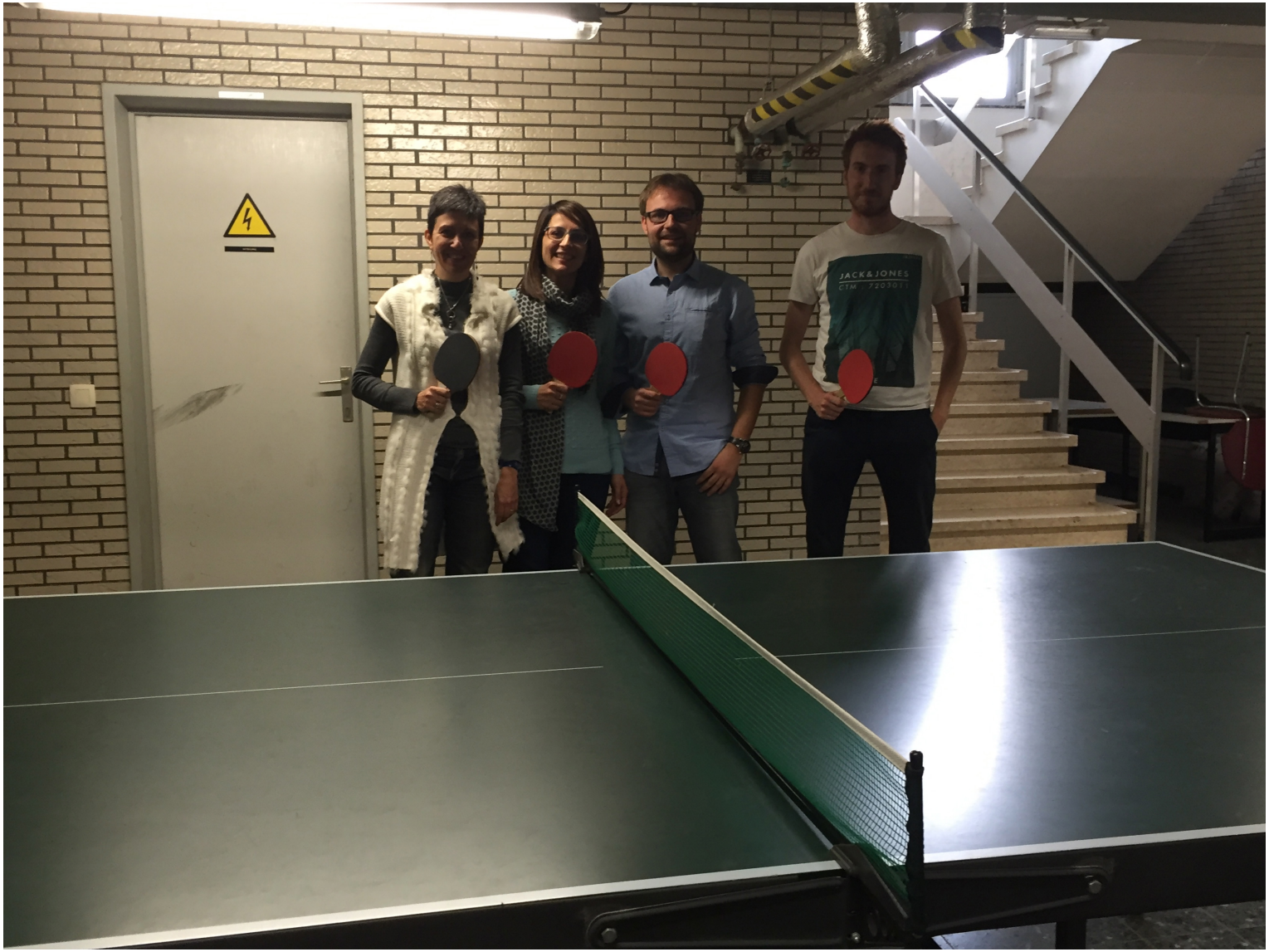
An Alternative Ranking for National Soccer Teams Based on Strength Parameters

Christophe Ley
Universiteit Gent

joint work with Hans Van Eetvelde (UGent) and Tom Van de Wiele (DeepMind)

BDSports: BIG DATA ANALYTICS IN SPORTS

November 16 2017


JACK&JONES

GABRIELS

An Alternative Ranking for National Soccer Teams Based on Strength Parameters

Christophe Ley
Universiteit Gent

joint work with Hans Van Eetvelde (UGent) and Tom Van de Wiele (DeepMind)

BDSports: BIG DATA ANALYTICS IN SPORTS

November 16 2017

Overview

1. The FIFA ranking
2. Various statistical models for football
3. Analysis of the English Premier League
4. New FIFA ranking and the EURO2016
5. Future research

What is the FIFA ranking ?

Shocking but great Belgian news on the 5th November 2015 ! ! !

How is it possible that Belgium is number one without winning any tournament ? What do the points mean ? How are they calculated ? And why is there such a FIFA ranking at all ?

The FIFA ranking is based on points of national team matches over the last 4 years !

The FIFA ranking is based on points of national team matches over the last 4 years !

The final points are calculated through the following formula :

Result $\times$Time $\times$Importance$\times$Opposition Strength$\times$Opposition Confederation.

More precisely :

- Result : the 3-1-0 system
- Time : 1 for the last 12 months, 0.5 for the months 13-24, 0.3 for months 25-36, 0.2 for months 37-48
- Importance : 4 for World Cup matches, 3 for EURO (or Africa Cup of Nations, ...), 2.5 for World Cup Qualifiers, 1 for friendly games
- Opposition Strength : max $(50, 200 -$ Opposition ranking$)$ with exception for number 1 (factor 200 instead of 199)
- Opposition Confederation : 1 for South-America, 0.99 for Europe, and 0.85 for other regions.

This ranking is suboptimal

The FIFA ranking is notorious for being bad and unfair in various aspects.

This ranking is suboptimal

The FIFA ranking is notorious for being bad and unfair in various aspects.

1. Not enough details are taken into account : goals, home or away,...

This ranking is suboptimal

The FIFA ranking is notorious for being bad and unfair in various aspects.

1. Not enough details are taken into account : goals, home or away,...
2. The Time effect : it is very abrupt ![1]

1. Scotland was ranked 50th in August 2013, ranked 63rd in September 2013, and then ranked 35th in October 2013.

This ranking is suboptimal

The FIFA ranking is notorious for being bad and unfair in various aspects.

1. Not enough details are taken into account : goals, home or away,...
2. The Time effect : it is very abrupt ![1]
3. The Opposition Strength formula max $(50, 200 -$ Opposition ranking$)$: just the rank is looked at, not the points of a team. Example : the difference between rank 31 (having 800 points) and rank 32 (799 points) is the same as the difference between rank 32 (799 points) and rank 33 (200 points)

1. Scotland was ranked 50th in August 2013, ranked 63rd in September 2013, and then ranked 35th in October 2013.

This ranking is suboptimal

The FIFA ranking is notorious for being bad and unfair in various aspects.

1. Not enough details are taken into account : goals, home or away,...
2. The Time effect : it is very abrupt ![1]
3. The Opposition Strength formula max $(50, 200 -$ Opposition ranking$)$: just the rank is looked at, not the points of a team. Example : the difference between rank 31 (having 800 points) and rank 32 (799 points) is the same as the difference between rank 32 (799 points) and rank 33 (200 points)
4. A team may “exploit” the system by avoiding playing friendly games who have less weight
5. The FIFA ranking does not lend itself to predictions of future matches

1. Scotland was ranked 50th in August 2013, ranked 63rd in September 2013, and then ranked 35th in October 2013.

This ranking is suboptimal

The FIFA ranking is notorious for being bad and unfair in various aspects.

1. Not enough details are taken into account : goals, home or away,...
2. The Time effect : it is very abrupt ![1]
3. The Opposition Strength formula max $(50, 200 -$ Opposition ranking$)$: just the rank is looked at, not the points of a team. Example : the difference between rank 31 (having 800 points) and rank 32 (799 points) is the same as the difference between rank 32 (799 points) and rank 33 (200 points)
4. A team may "exploit" the system by avoiding playing friendly games who have less weight
5. The FIFA ranking does not lend itself to predictions of future matches

The Importance-factor is also often criticized, but how to define it differently ?

1. Scotland was ranked 50th in August 2013, ranked 63rd in September 2013, and then ranked 35th in October 2013.

The Time effect

This issue can easily be solved : replace the step-function with a continuously decaying function.

The Time effect

This issue can easily be solved : replace the step-function with a continuously decaying function. Our proposal

$$\text{Time effect for } x_i \text{ days} = \exp\left(-\frac{x_i}{\text{Half period}}\log(2)\right).$$

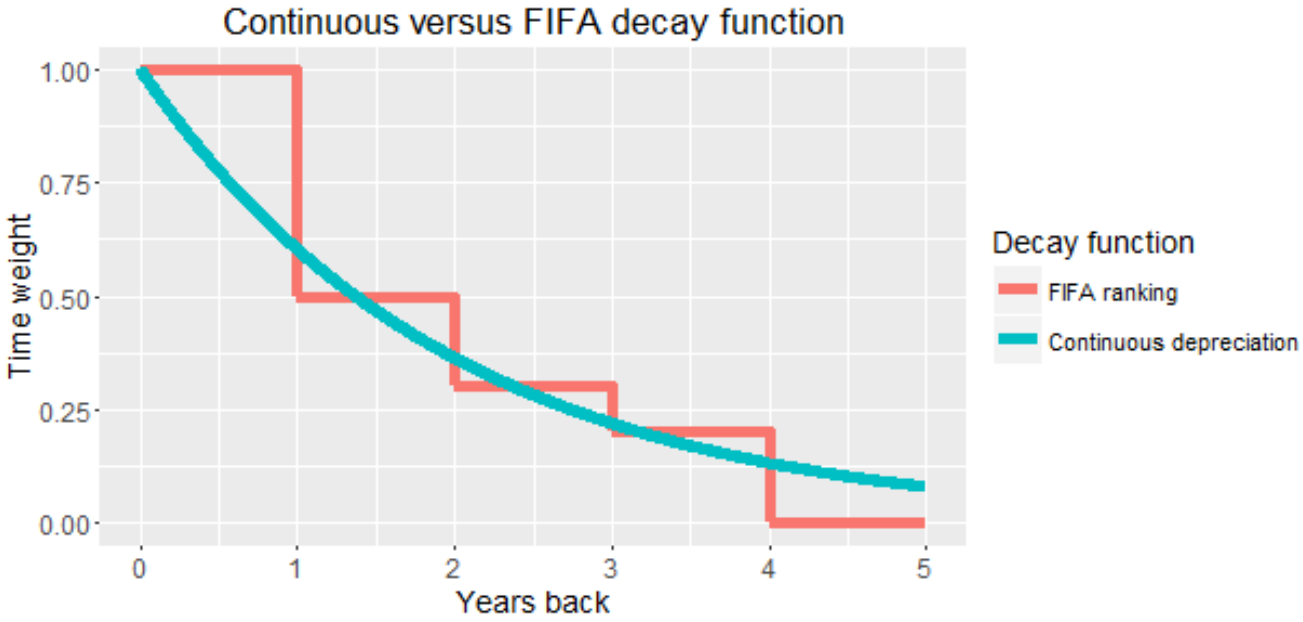


FIGURE : For a half period of 500 days

Overview

1. The FIFA ranking
2. Various statistical models for football
3. Analysis of the English Premier League
4. New FIFA ranking and the EURO2016
5. Future research

The Bradley-Terry type model

Each of M teams is assigned a positive team strength β_j. The outcome of match $i \in \{1, \ldots, N\}$ is then modelled as

$$P_{iH} = \frac{h\beta_{iH}}{h\beta_{iH} + d\sqrt{h\beta_{iH}\beta_{iA}} + \beta_{iA}};$$
$$P_{iD} = \frac{d\sqrt{h\beta_{iH}\beta_{iA}}}{h\beta_{iH} + d\sqrt{h\beta_{iH}\beta_{iA}} + \beta_{iA}};$$
$$P_{iA} = \frac{\beta_{iA}}{h\beta_{iH} + d\sqrt{h\beta_{iH}\beta_{iA}} + \beta_{iA}},$$

with home effect represented by h and draw effect by d; h is dropped on neutral grounds.

This model was used by Wang and Vandebroek (2011); the multiplicative home effect is based on an analysis of 10 Serie A seasons.

These strength parameters are estimated by maximum likelihood :

$$L = \prod_{i=1}^{N} \prod_{j \in \{H,D,A\}} P_{ij}^{y_{ij} \cdot w_{type,i} \cdot w_{time,i}}$$

with

- y_{ij} an indicator function of the result
- $w_{type,i}$ the match importance factor (we keep the FIFA weights)
- $w_{time,i}$ the time depreciation effect

The Extended Bradley-Terry model

A limitation of the BT model : same home effect parameter h !

The Extended Bradley-Terry model

A limitation of the BT model : same home effect parameter h !

The EBT model assigns two strength parameters to each team, one for home and one for away.

How to combine the home and away strengths ? Via geometric mean (better than arithmetic mean and harmonic mean according to an analysis of all Premier League data from 1892-2014).

The Extended Bradley-Terry model

A limitation of the BT model : same home effect parameter h !

The EBT model assigns two strength parameters to each team, one for home and one for away.

How to combine the home and away strengths ? Via geometric mean (better than arithmetic mean and harmonic mean according to an analysis of all Premier League data from 1892-2014).

The parameters are again ML-estimated.

The Poisson model

Maher (1982) assumed the number of scored goals by both teams (G_{iH} and G_{iA}) to be independent Poisson distributed variables :

$$\mathrm{P}(G_{iH} = x, G_{iA} = y) = \frac{\lambda_{iH}^{x}}{x!} \exp(-\lambda_{iH}) \cdot \frac{\lambda_{iA}^{y}}{y!} \exp(-\lambda_{iA}),$$

where λ_{iH} and λ_{iA} are the means of the scored goals for the home and the away team in match i.

With strength parameters β_{iH} and β_{iA}, we consider $\lambda_{iH} = c \times h \times \frac{\beta_{iH}}{\beta_{iA}}$ and $\lambda_{iA} = c \times \frac{\beta_{iA}}{\beta_{iH}}$ under the constraint that all parameters are positive ; c is a common intercept. Matches on neutral ground are modeled by dropping the home effect h.

ML-estimation :

$$\prod_{i=1}^{N} \left(\frac{\lambda_{iH}^{g_{iH}}}{g_{iH}!} \exp(-\lambda_{iH}) \cdot \frac{\lambda_{iA}^{g_{iA}}}{g_{iA}!} \exp(-\lambda_{iA}) \right)^{\omega_{time,i} \times \omega_{type,i}}$$

The Poisson model allows easy computation of the probabilities P_{ij}. Writing $D = G_{iH} - G_{iA}$, we look at $\mathrm{P}(D > 0)$, $\mathrm{P}(D = 0)$ and $\mathrm{P}(D < 0)$. The distribution of D even bears a name : the Skellam distribution.

Compared to BT and ETB, the Poisson model uses two observations per match instead of one.

The Extended Poisson model

No home-away change, but offensive (o_{iH}, o_{iA}) and defensive (d_{iH}, d_{iA}) strengths for each team !

Thus $\lambda_{iH} = c \times h \times \frac{o_{iH}}{d_{iA}}$ and $\lambda_{iA} = c \times \frac{o_{iA}}{d_{iH}}$.

This is the original Maher (1982) model.

The Combined Bradley-Terry - Poisson model

Idea : augment the BT model through a goal difference effect.

$$L = \prod_{i=1}^{N} \prod_{j \in \{H,D,A\}} P_{ij}^{y_{ij} \cdot w_{goalDiffscaled,i} \cdot w_{type,i} \cdot w_{time,i}}$$

with

$$w_{goalDiffscaled,i} = \begin{cases} 1 & \text{if draw} \\ \log_2(goalDiff_i + 1) & \text{else} \end{cases}$$

with $goalDiff_i$ the (absolute) goal difference in match i.

- Goal diff of 1 $\Longrightarrow$ weight 1
- Goal diff of 7 $\Longrightarrow$ weight 3

The ELO model

Used in chess and female FIFA ranking !

The ELO model

Used in chess and female FIFA ranking !

Difference with other models : strength parameters are evaluated sequentially, after each match $\Rightarrow$ naturally more weight on more recent matches.

The ELO model

Used in chess and female FIFA ranking !

Difference with other models : strength parameters are evaluated sequentially, after each match $\Rightarrow$ naturally more weight on more recent matches. Update formula :

$$R_n = R_o + k \cdot (O_{act} - O_{exp})$$

where R_n is the new rating, R_o the old rating, k the match importance, O_{act} the actual outcome and O_{exp} the expected match outcome.

Calculation of O_{exp} :

$$O_{exp} = \frac{1}{1 + 10^{-\frac{R_o - A_o + h}{400}}}$$

with A_o the opponent rating before the match and h the home effect (100 for home team, -100 for away team, 0 on neutral ground). The match importance factor k is proportional to those from the male FIFA ranking.

Calculation of O_{act} :

TABLE : Percentage scores of O_{act} for the **drawn or losing team**. The winner is awarded 1-$O_{act,loser}$, except for a draw (goal difference = 0) when the opponent receives the same number of points.

	Goal difference						
	0	1	2	3	4	5	6/6+
Goals scored	**Actual result (percentage)**						
0	47	15	8	4	3	2	1
1	50	16	8.9	4.8	3.7	2.6	1.5
2	51	17	9.8	5.6	4.4	3.2	2
3	52	18	10.7	6.4	5.1	3.8	2.5
4	52.5	19	11.6	7.2	5.8	4.4	3
5/5+	53	20	12.5	8	6.5	5	3.5

Overview

1. The FIFA ranking
2. Various statistical models for football
3. Analysis of the English Premier League
4. New FIFA ranking and the EURO2016
5. Future research

Premier League

We compare the 6 models on basis of the Premier League seasons between 2000-2001 and 2014-2015 by measuring their predictive performance via the log loss :

$$log\ loss = -\frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{3} y_{ij} \cdot log(P_{ij})$$

where P_{ij} represents the modeled probabilities and y_{ij} the actual outcome of the match.

Prediction based on 2850 matches. We make a comparison with the bookmaker data model and the match data benchmark model.

An alternative performance indicator is the Rank Probability Score (RPS ; difference between cumulative predicted and observed distribution) from Epstein (1969).

Bookmaker data

The website `http://www.football-data.co.uk/englandm.php/` hosts a separate excel file with match data and odds of each English Premier League season since the season 2000-2001.

If the observed odds are [2,3,3.5] for a home win, a draw and a home loss respectively, the transformed probabilities become $[\frac{1}{2}, \frac{1}{3}, \frac{1}{3.5}]/(\frac{1}{2} + \frac{1}{3} + \frac{1}{3.5}) \approx [0.447, 0.298, 0.255]$.

The benchmark - match data

These data are designed for post-hoc analysis. They contain the difference in total shots, shots on target, corners, fouls as well as yellow and red cards.

Most important predictor : difference in shots on target.

Additional shots off target reduce the winning chances. Idem for number of corners !

Cards have no significant influence.

Results of the comparison (1)

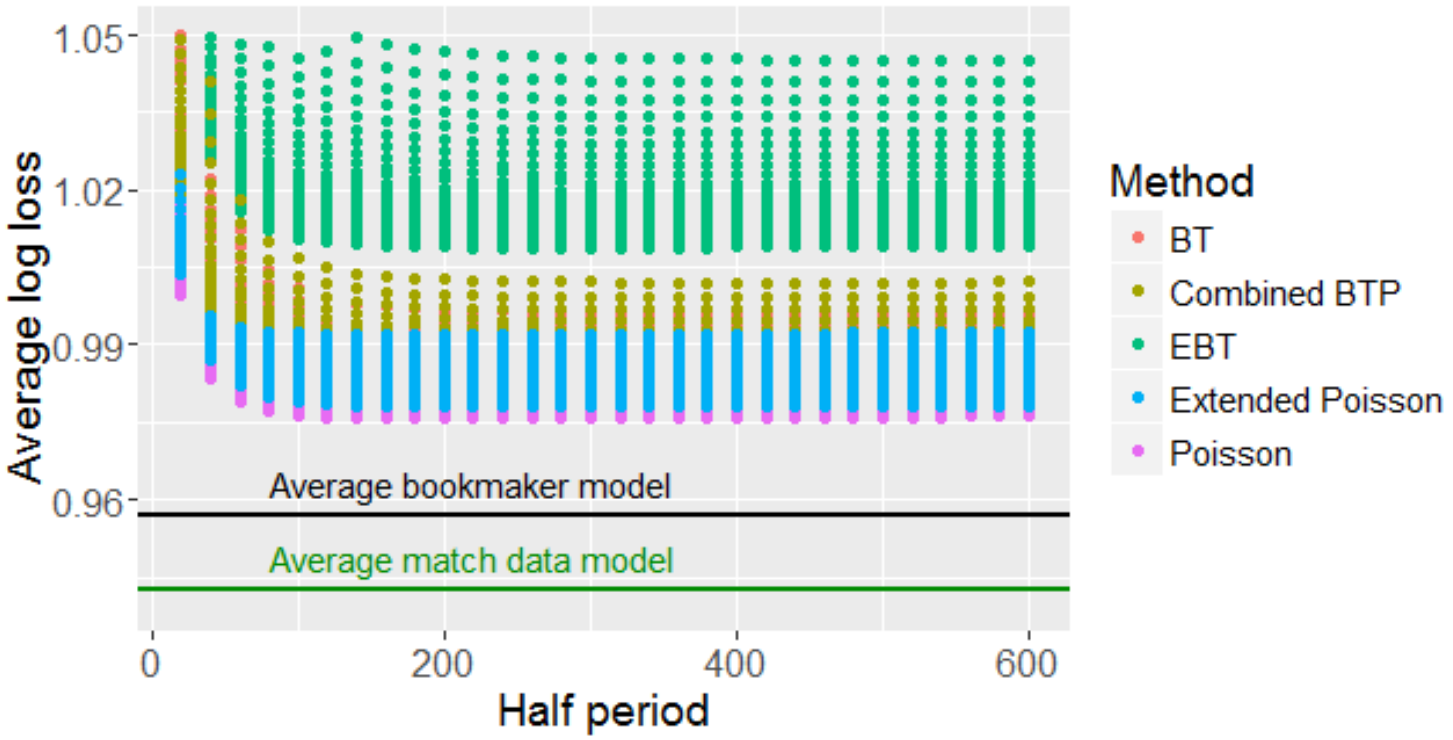


FIGURE : Model comparison graph using all 2850 second season half English Premier League matches in the period between the 2000-2001 and the 2014-2015 season. Each model is fit using different half periods and relaxation parameters. The model with the lowest log loss value is a Poisson model with a half period of 180 days.

Results of the comparison (2)

TABLE : Comparison table for the best performing models of each of the six considered classes with respect to the average log loss criterion and two benchmark models.

Model Class	**Lowest log loss**	**Model Class Rank**	**Time depreciation constant best model**	**Correlation match statistics model**	**Correlation average bookmakers model**
BT	0.986	6	HP = 340	0.459	0.913
Poisson	0.976	3	HP = 180	0.501	0.941
EBT	1.009	8	HP = 280	0.406	0.810
Extended Poisson	0.978	4	HP = 300	0.492	0.931
Combined BTP	0.985	5	HP = 220	0.449	0.918
Elo	0.991	7	Elo const = 26	0.495	0.878
Match statistics	0.942	1	/	1	0.534
Bookmakers avg	0.957	2	/	0.534	1

Focus on Poisson versus Bookmakers

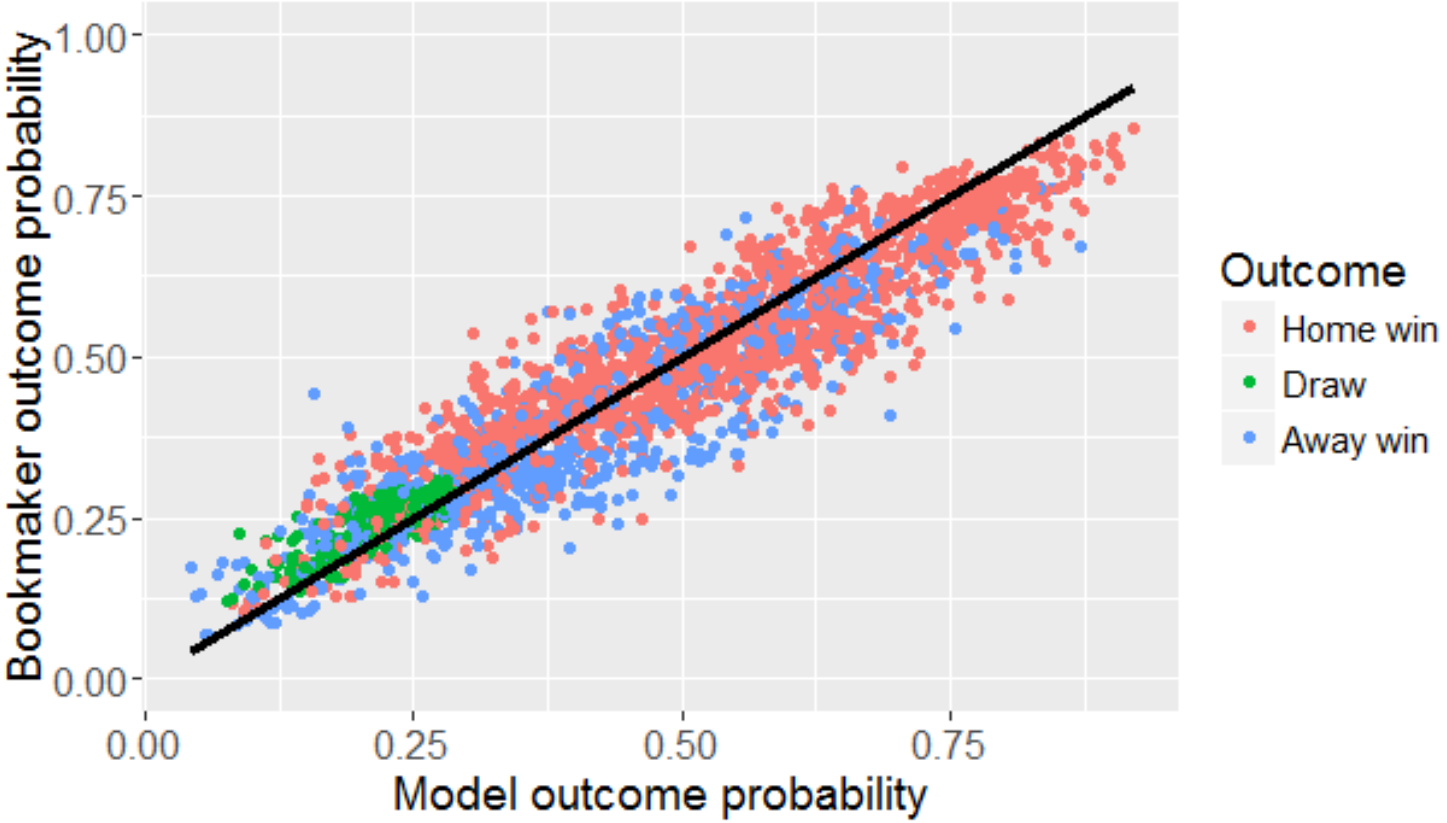


FIGURE : Comparison of the average bookmaker actual outcome probability and the actual outcome probability for the Poisson model with half period 180 days. The Pearson linear correlation coefficient is 0.941.

Overview

1. The FIFA ranking
2. Various statistical models for football
3. Analysis of the English Premier League
4. New FIFA ranking and the EURO2016
5. Future research

National Team data

National team match results were scraped from the website http ://eu-football.info.

The platform contains a complete archive of all European national football team results (1872-2016) where at least one of the playing teams was European. The website also contains historical results of all European and domestic club competitions.

The Poisson model was again found to be the best model in terms of log-loss !

It was trained on 3785 national team matches (taking out countries with <100.000 inhabitants) and yielded a half-period of 1320 days. Only matches with two European participants were considered.

UEFA Euro 2016

Portugal was the big winner, but this came rather as a surprise as they also did not play the best football.

UEFA
EURO2016
FRANCE
POL 1 - 1P POR
GER 1P - 1 ITA
1
0
2 - 0
0 - 2
WAL 3 - 1 BEL
FRA 5 - 2 ISL

Team	Model rank	FIFA rank	Team	Model rank	FIFA rank
Germany	1	2 (4)	Hungary	28	13 (20)
France	2	10 (17)	Wales	29	18 (26)
Spain	3	3 (6)	Northern Ireland	30	17 (15)
Belgium	4	1 (2)	Israel	31	35 (71)
England	5	6 (11)	Norway	32	30 (51)
Netherlands	6	8 (14)	Greece	33	27 (40)
Ukraine	7	12 (19)	Bulgaria	34	34 (69)
Portugal	8	4 (8)	Finland	35	33 (67)
Russia	9	21 (29)	Belarus	36	36 (78)
Austria	10	5 (10)	Montenegro	37	39 (90)
Croatia	11	19 (27)	Cyprus	38	37 (84)
Poland	12	20 (27)	Macedonia	39	47 (139)
Italy	13	7 (12)	Georgia	40	45 (137)
Bosnia & Herzegovina	14	14 (20)	Armenia	41	42 (110)
Turkey	15	11 (18)	Kazakhstan	42	43 (112)
Switzerland	16	9 (15)	Latvia	43	41 (104)
Czech Republic	17	22 (30)	Moldova	44	49 (159)
Sweden	18	25 (35)	Azerbaijan	45	46 (138)
Serbia	19	31 (54)	Estonia	45	40 (94)
Iceland	20	24 (34)	Lithuania	47	44 (127)
Romania	21	15 (22)	Faroe Islands	48	38 (89)
Republic of Ireland	22	23 (33)	Luxembourg	49	48 (146)
Slovakia	23	16 (24)	Malta	50	50 (166)
Scotland	24	29 (43)	Liechtenstein	51	51 (168)
Denmark	25	26 (38)	Andorra	52	53 (202)
Albania	26	28 (42)	San Marino	53	52 (200)

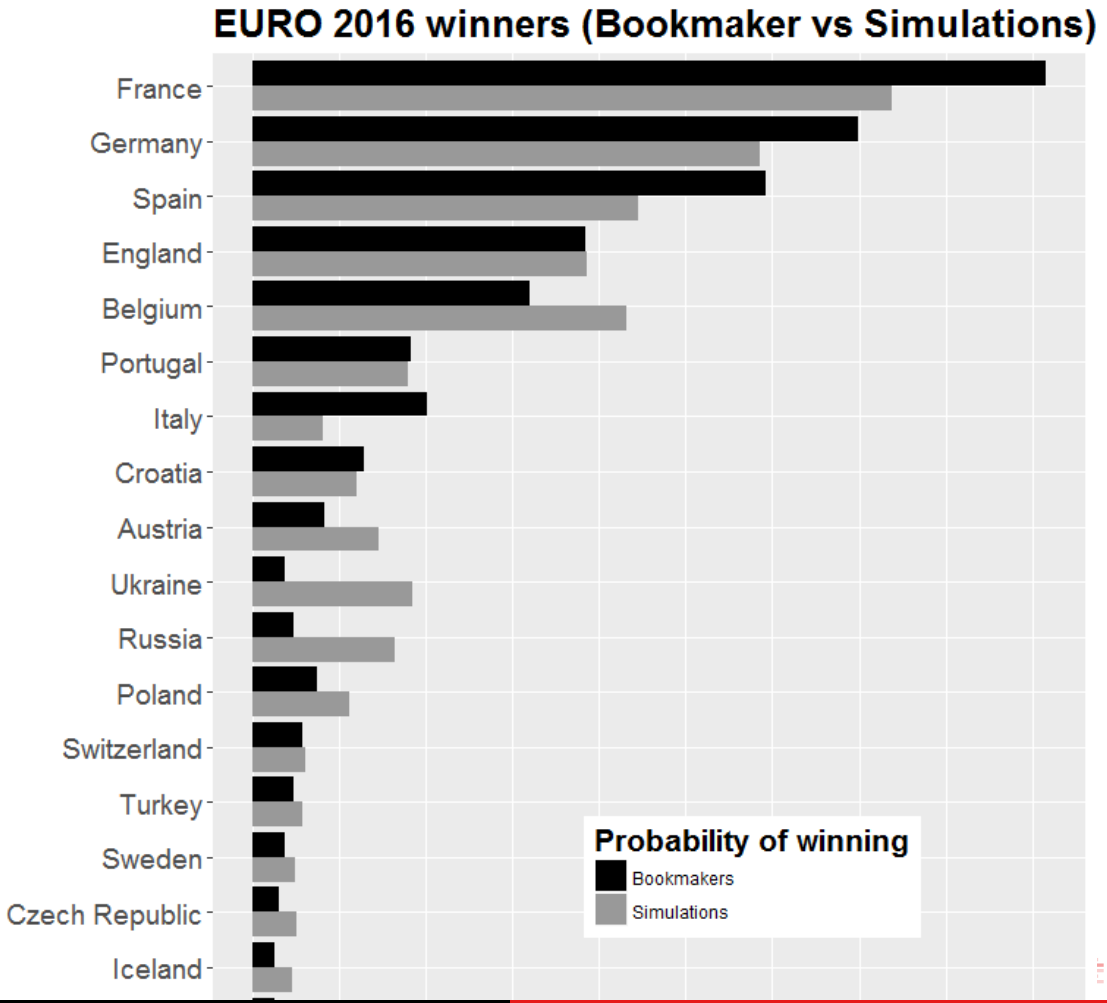
(Post-)Predicting the first matches of the EURO2016

TABLE : We compare the actual results with the results that have the greatest probability according to our model.

Game	Actual outcome	Model outcome	Model probabilities actual Outcome
France - Romania	2 - 1	1 - 0	0.094
Albania - Switzerland	0 - 1	0 - 1	0.135
Wales - Slovakia	2 - 1	0 - 1	0.061
England - Russia	1 - 1	1 - 0	0.130
Turkey - Croatia	0 - 1	0 - 1	0.135
Poland - Northern Ireland	1 - 0	1 - 0	0.152
Germany - Ukraine	2 - 0	1 - 0	0.100
Spain - Czech Republic	1 - 0	1 - 0	0.153
Republic of Ireland - Sweden	1 - 1	1 - 1	0.134
Belgium - Italy	0 - 2	1 - 0	0.033
Austria - Hungary	0 - 2	1 - 0	0.028
Portugal - Iceland	1 - 1	1 - 0	0.128

(Post-)Predicting the entire EURO2016

We have simulated 100.000 times the UEFA Euro 2016. The correlation between simulated and bookmaker ranks is $\rho = 0.893$.



Overview

1. The FIFA ranking
2. Various statistical models for football
3. Analysis of the English Premier League
4. New FIFA ranking and the EURO2016
5. Future research

Much remains to be done

- For the FIFA ranking : are the weights inherited from the official FIFA ranking really the best choices ? What about the continental weights ?

Much remains to be done

- For the FIFA ranking : are the weights inherited from the official FIFA ranking really the best choices ? What about the continental weights ?
- Compare to dependent Poisson models and Copula-based models

Much remains to be done

- For the FIFA ranking : are the weights inherited from the official FIFA ranking really the best choices ? What about the continental weights ?
- Compare to dependent Poisson models and Copula-based models
- Compare to Machine Learning prediction models (cf. the competition of the journal Machine Learning)

Much remains to be done

- For the FIFA ranking : are the weights inherited from the official FIFA ranking really the best choices ? What about the continental weights ?
- Compare to dependent Poisson models and Copula-based models
- Compare to Machine Learning prediction models (cf. the competition of the journal Machine Learning)
- Take into account clever chosen covariates (see next talk)

Grazie mille !