KU LEUVEN

FACULTY OF ECONOMICS
AND BUSINESS

A model based ranking system for soccer teams

Chang Wang, Martina Vandebroek

KBI_1307

A Model Based Ranking System for Soccer Teams

Chang Wang

Faculty of Economics and Business, KU Leuven, Belgium

Martina Vandebroek

Faculty of Economics and Business and Leuven Statistics Research Center, KU Leuven, Belgium

Summary. In competitive sports, the design of an effective ranking system attracts increasing interests. For sports with 2-way results, the Bradley-Terry model has proven to be an effective approach. However, a well-known limitation of standard Bradley-Terry models is that they cannot be applied to sports which allow to end with a tie after regular playing time. During the past decades various methods have been explored to model soccer games in the literature. In this paper, a new extended Bradley-Terry model with various covariates is devised to construct a more advanced ranking system for international soccer teams based on their games played within the past 4 years. First a standard likelihood (ML) estimation method was employed. A weighted likelihood (WML) estimation procedure was run for further improvement. Finally an evolutional model was developed to capture the dynamic trend in performance during big events. This paper examines the extent to which our ranking differs from the existing FIFA ranking. In addition, the 2012 UEFA European Football Championships was analyzed.

Keywords: Bradley-Terry model; Evolutionary model; Multinomial distribution; Strength of soccer teams; Weighted likelihood

1. Introduction

In many competitive sports, paired comparisons are used for constructing a ranking system. Stefani (1997) presented a survey of the major world sports rating systems, including tennis and soccer. He summarized the 3 common phases for each rating system, which can be labeled as weighting, summation and averaging. During the past decades, statistical models for ordered paired comparisons have been proposed for more advanced evaluation of competitions.

One of the most prominent and popular models for paired comparisons was devised by Bradley and Terry (1952). Since its introduction, the standard Bradley-Terry model has received extensive attention in the literature. Although the Bradley-Terry model has proven to be the standard approach for 2-way outcomes, it does not allow to model ties. In sports, ties occur when there is no difference in score at the end of regular playing time. More general, ties exist when a judge is allowed a third opinion, that of no preference (Dittrich et al., 1998). In the literature the usual strategy is either to force a definite expression of preference, or by ignoring or random allocating the occurring ties (Davidson 1970). In order to overcome this limitation in modeling 3-way outcome comparison, various methods have been developed in the literature. These approaches can be categorized into two main streams.

Address for correspondence: KU Leuven, Naamsestraat 69, B-3000 Leuven.
E-mail: chang.wang@kuleuven.be

The first class of approaches apply a linear model to the difference in score from each game. Then a tie will be declared when the absolute difference between two opponents lies below a certain threshold. Stern (1992) used least squares to fit a threshold model to generate a football ranking. Bassett (1997) argued that using sports rating based on least absolute errors is more reliable than ratings based on the least squares approach. Dyte and Clarke (2000) suggested that the number of goals scored by each team in a game was independently Poisson distributed, with mean depending on the FIFA points of both teams and the match venue. Although these methods are innovative, the selection of the threshold to be used while estimating a continuous response is rather subjective.

The second class of approaches analyze the data by considering only the outcomes (win-tie-loss) of the matches. Glenn and David (1960) used a modified Thurstone-Mosteller type model to provide for tied observations. The difference between the performances of two opponents was assumed normally distributed. Rao and Kupper (1967) replaced the normal distribution in the Glenn-David model by the logistic distribution. Davidson (1970) proposed an extension to the Bradley-Terry model to handle tied games. Fahrmeir and Tutz (1994) devised a dynamic stochastic model which consists of two components: a response model which connects the observed results and the underlying team abilities, and a transition model which specifies the variation of team abilities over time. Kuk (1995) proposed a linear paired comparison model with home and away strength parameters for each team.

Besides these two categories, Stern (2007) investigated situations where the observed measure of preference for the paired comparisons is instead a continuous outcome indicating not only the outcome direction but also the degree of preference. His new method, which is labeled *moderated paired comparisons*, is based on fitting a penalized likelihood model to the margin of victory.

In this study, a new ranking system for soccer teams is proposed which takes into account the home effect. For this we start from the extended Bradley-Terry model that was developed by Davidson (1970) to allow three-way outcomes. The model is estimated based on a data set which is collected from FIFA's official website. In order to keep the computational complexity at a feasible level, we only consider games between two UEFA teams. Our data set involves all 53 UEFA international teams' 929 official games played within the time interval from January 1^{st}, 2008 until December 31^{st}, 2011. First standard maximum likelihood (ML) will be used to fit the model, then a weighted ML version, taking into account the time and game type impact, is used. Finally an innovative evolutional model is devised to capture the evolution of a teams' performance in the final round of tournaments. The paper also examines the extent to which our model-based ranking is better than the points-based method in terms of predictive power.

The paper is organized as follows. Section 2 gives a brief overview of the FIFA ranking. Section 3 presents an extended Bradley-Terry model taking ties and the home advantage into account. The model is fitted in Section 4 to the data for UEFA countries using unweighted and weighted maximum likelihood. Section 5 describes an evolutional version of the proposed model and compares the results with previous versions. Section 6 applies our final model to analyze the UEFA EURO 2012 and discusses its forecasting ability. Concluding remarks and plans for future research are summarized in Section 7.

2. The Coca-Cola FIFA ranking

The current Coca-Cola FIFA world ranking (FIFA, 2011) is based on the calculation of points. Any team that performs well in international football games earns points and improves its ranking position. One team's total number of points over a four-year period is determined by adding the points within four 12-month windows. There is a depreciation for the matches older than 12 months. All games played within a 12-month period are depreciated similarly. The detailed Coca-Cola FIFA world ranking procedure is referred to on the official FIFA website. Here we only give a brief overview of its methodology.

2.1. The FIFA point system

The calculation of the points for a single match is simple as it is the product of the following four values:

M stands for the points for the match result. Teams gain 3 points for a victory, 1 point for a draw and 0 points for a defeat. In a penalty shoot-out, the winning team gains 2 points and the losing team gains 1 point.

I stands for importance of the match. This value gives the relative importance of different types of matches:

Friendly match (including small competitions)	1.0
FIFA World Cup qualifier or confederation-level qualifier	2.5
Confederation-level final competition or FIFA Confederations Cup	3.0
FIFA World Cup final competition	4.0

T stands for the strength of the opposing team. This value is calculated as *200 minus the ranking position of the opponent in the most recently published Coca-Cola FIFA World Ranking*. As an exception, the 1^{st} ranked team is always assigned the value 200 and the teams ranked 150^{th} and below are assigned a minimum value of 50.

C stands for the strength of the confederation. This value is decided by the confederation to which the opponent team belongs. The strength coefficients of the confederations are as follows:

UEFA/CONMEBOL	1.00
CONCACAF	0.88
AFC/CAF	0.86
OFC	0.85

The teams final number of points is calculated after each month by a weighted sum of the average points in the previous 4 years. The weights depreciate older matches as follows:

Match average from the past 12 months	1.0
Match average from the past 24 to 12 months	0.5
Match average from the past 36 to 24 months	0.3
Match average from the past 48 to 36 months	0.2

The first three columns in table 1 give the FIFA points and the corresponding ranking on November 23rd, 2011 for the 53 teams considered in this study.

Table 1. Comparison of the FIFA ranking and the results of the 3 models developed in this paper together with the estimated parameters

	FIFA		ML		WML		EVM	
Country	points	rank	strength	rank	strength	rank	strength	rank
Spain	1564	1	113.069	1	92.595	1	79.294	1
Netherlands	1365	2	36.162	2	29.253	2	26.069	2
Germany	1345	3	19.688	4	22.209	3	21.124	3
England	1173	4	25.560	3	15.999	4	14.950	4
Portugal	1100	5	6.980	8	9.034	6	8.721	6
Croatia	1091	6	9.412	6	6.693	9	6.660	10
Italy	1082	7	13.943	5	10.123	5	9.915	5
Denmark	1035	8	5.328	12	8.007	8	7.740	8
Russia	971	9	6.767	9	6.441	10	7.126	9
Greece	964	10	4.104	14	4.909	11	4.896	11
France	915	11	7.950	7	4.802	12	4.763	12
Switzerland	898	12	3.083	17	3.343	15	3.412	15
Sweden	891	13	5.646	10	8.107	7	7.783	7
Republic of Ireland	864	14	3.425	15	3.484	14	3.581	14
Bosnia-Herzegovina	838	15	2.732	18	2.984	17	2.972	17
Norway	788	16	3.287	16	3.220	16	3.176	16
Slovenia	780	17	0.866	30	0.924	28	0.947	28
Serbia	775	18	2.555	19	2.403	19	2.538	19
Turkey	769	19	4.406	13	2.460	18	2.553	18
Czech Republic	749	20	1.765	20	2.224	20	2.230	20
Hungary	665	21	1.559	22	2.219	21	2.170	21
Israel	665	22	1.235	26	0.998	27	1.001	26
Slovakia	648	23	0.654	33	1.015	25	1.093	25
Belgium	647	24	1.415	24	1.202	23	1.213	23
Armenia	612	25	0.319	40	0.534	34	0.552	34
Scotland	599	26	1.522	23	1.682	22	1.659	22
Wales	591	27	1.000	29	1.000	26	1.000	27
Montenegro	585	28	1.143	28	0.693	31	0.693	31
Ukraine	568	29	5.607	11	3.810	13	3.770	13
Romania	562	30	1.649	21	1.116	24	1.134	24
Estonia	543	31	0.376	39	0.648	33	0.659	33
Belarus	501	32	0.860	31	0.660	32	0.660	32
Poland	487	33	1.413	25	0.795	29	0.804	29
Latvia	466	34	0.616	34	0.469	36	0.470	36
Austria	459	35	0.542	35	0.453	37	0.458	37
Georgia	451	36	0.256	42	0.209	42	0.209	42
Albania	450	37	0.427	38	0.268	38	0.269	38
Bulgaria	410	38	1.146	27	0.493	35	0.494	35
Finland	405	39	0.817	32	0.720	30	0.711	30
Northern Ireland	403	40	0.250	43	0.234	40	0.240	40
Lithuania	382	41	0.461	37	0.250	39	0.252	39
FYR Macedonia	333	42	0.464	36	0.225	41	0.231	41
Iceland	325	43	0.235	44	0.180	43	0.179	43
Azerbaijan	300	44	0.149	46	0.112	45	0.119	45
Faroe Islands	285	45	0.069	47	0.064	48	0.065	48
Cyprus	253	46	0.290	41	0.136	44	0.135	44
Luxembourg	252	47	0.062	48	0.060	49	0.060	49
Liechtenstein	246	48	0.046	50	0.076	47	0.076	47
Moldova	229	49	0.155	45	0.080	46	0.080	46
Kazakhstan	219	50	0.049	49	0.043	50	0.044	50
Malta	140	51	0.043	51	0.021	51	0.021	51
Andorra	0	52	0.000	52	0.000	52	0.000	52
San Marino	0	52	0.000	52	0.000	52	0.000	52
Home effect (H)	-	-	2.080	-	2.177	-	2.167	-
Tie effect (K)	-	-	0.905	-	0.893	-	0.895	-
Stage effect (G)	-	-	-	-	-	-	1.430	-

2.2. Some issues about the FIFA ranking

The FIFA ranking does not only evaluate the relative strength of all 209 international teams, but is also useful when categorizing the teams within a continent into different pots. This is especially helpful for the draws in the FIFA World Cup qualifiers or continental championships. Therefore, every team can benefit from improving its position in the FIFA ranking.

Although the existing FIFA ranking is used for different purposes, there are several drawbacks to be noted. Firstly, it is a points-earning system. Theoretically speaking, two teams both ranked 150^{th} or below (e.g., Andorra and San Marino, sharing a 206^{th} position and 0 points) can always benefit from playing a large number of friendly games against each other and negotiate to win half of these games each. They can apply this strategy without any cost, because neither will receive negative points when losing a game. Secondly, the current ranking assigns a fixed depreciation weight for all games played within a 12-month period. Although its basic idea is that the older a game was, the lower weight should be assigned to it, its depreciation function is piecewise with 3 discontinuous points, which seems not very reasonable. As such, there is clearly some room for improving the time impact in the ranking. Thirdly, although the points-based FIFA ranking is easy to operate and modify, it is not obvious how the points should be used to predict the 3 possible outcomes in a game. Lastly, the FIFA ranking does not distinguish the home and away team in a game. As the average performance difference between home and away games for almost all football teams has been revealed by various researchers, a team can benefit more from playing a higher percentage of its games at home. A more realistic method should then take the home effect into account such that the ranking can be determined by the pure strength parameters after the home effect is removed. In short, a new ranking method is called for to overcome the drawbacks of the FIFA ranking. Instead of the points-based FIFA ranking system, we will design a ranking based on statistic modeling.

3. A model-based approach

A regular soccer game consists of two 45-minute halves. It ends after the regular time of 90 minutes for almost all A-International friendly matches, unofficial championship matches, and qualifier matches of official tournaments. In the final round of the FIFA World Cup or continental championship tournaments, a group stage game ends after 90 minutes regardless whether there is a goal difference or not. In the knockout stage, a game ends after regular playing time if there is a goal difference. Otherwise a 30-minute extra time will be added. If after the extra time there is still no goal difference, a penalty shootout will be played.

3.1. The 3-way probabilistic model

In this paper, a new model-based instead of points-based ranking will be derived. As mentioned previously, we need to model the probability of ties in a soccer match which is not available in the standard Bradley-Terry model. There are various ways for constructing an extended Bradley-Terry models in the literature. We will model the tie probability similarly to Davidson (1970). The basic idea is driven by regularly occurring patterns in the football world. The closer in strength the two playing teams are, the more likely a tie is. In contrast, if there is a relatively huge strength difference between the two playing teams, it's very likely that the stronger team will win the game. With this in mind, we model P_{i1}, the probability that team 1 wins game i, P_{i2}, the probability that game i ends with a tie and P_{i3}, the probability that team 2 wins game i as follows

$$\begin{aligned} P_{i1} &= \frac{\beta_{1i}}{\beta_{1i}+\beta_{2i}+K\sqrt{\beta_{1i}\beta_{2i}}}, \\ P_{i2} &= \frac{K\sqrt{\beta_{1i}\beta_{2i}}}{\beta_{1i}+\beta_{2i}+K\sqrt{\beta_{1i}\beta_{2i}}}, \\ P_{i3} &= \frac{\beta_{2i}}{\beta_{1i}+\beta_{2i}+K\sqrt{\beta_{1i}\beta_{2i}}} \end{aligned} \tag{1}$$

where β_{1i} and β_{2i} denote the strength parameters of the first and second team in game i respectively. To ensure positive β_i values we replace them by $\exp(\alpha_i)$ in the estimation procedure. K is the tie effect to be estimated.

3.2. The 3-way probabilistic model including a home effect

In order to make our model more realistic, we will also take the home advantage into account. In soccer games, a team's performance usually varies considerably between home and away matches. This can be due to either natural factors or anthropological factors. Natural factors include the adeptness to the grass field, temperature, humidity, lighting, and other environmental effects. While anthropological factors refer to the support from the home side fans, that will dominate the number of spectators in a game. Usually they will put the referees under pressure, forcing them to make some favorable decisions for the home side. In an analysis of 10 recent Italian Serie A seasons it was apparent that the stronger teams have a larger home advantage than weaker teams. This effect is also noticeable in other influential domestic leagues. Therefore, we assume that the home advantage is proportional to the strength of the home team, instead of using an additive constant as in Bassett (1997). From now on we denote by team 1 the home team and by team 2 the away team. When the home advantage effect H is included, the model becomes

$$\begin{aligned} P_{i1} &= \frac{H\beta_{1i}}{H\beta_{1i}+\beta_{2i}+K\sqrt{H\beta_{1i}\beta_{2i}}}, \\ P_{i2} &= \frac{K\sqrt{H\beta_{1i}\beta_{2i}}}{H\beta_{1i}+\beta_{2i}+K\sqrt{H\beta_{1i}\beta_{2i}}}, \\ P_{i3} &= \frac{\beta_{2i}}{H\beta_{1i}+\beta_{2i}+K\sqrt{H\beta_{1i}\beta_{2i}}}. \end{aligned} \tag{2}$$

If the game is played on a neutral field, which is common in the final round of official tournaments, the home advantage will be eliminated when fitting the model by putting H equal to 1.

4. Estimation methods

4.1. Maximum likelihood estimation

We first fit the model using maximum likelihood, giving all games the same weight for estimating the team strength parameters. Let y_{ij} be 1 if the result of game i is j and $y_{ij} = 0$ otherwise, with $j = 1$ indicating that the home team won, $j = 2$ denoting a tie and

$j = 3$ that the away team won. Then the likelihood becomes

$$L = \prod_{i=1}^{N} \prod_{j=1}^{3} P_{ij}^{y_{ij}}. \tag{3}$$

The parameter estimates maximize the logarithm of this likelihood function which is highly non-linear. We used the R-functions *nlm* as well as *optim* to carry out the optimization. To ensure a unique solution and without loss of generality we choose Wales, which is alphabetically the last of the 53 teams considered, as the benchmark team and set its strength equal to 1 ($\beta_{53} = 1$).

Column 4 of Table 1 reports the estimated parameters for all 53 team's strengths, the home advantage as well as the tie effect. The corresponding ranking positions are given in column 5. All the estimated parameters converged well as the gradients were close to 0. Different starting values were used, and the estimated parameters turned out to be rather robust with respect to the starting values given a specified iteration limit of 250.

The estimated tie effect is 0.905. As the tie effect is smaller than 1, this indicates that if two equally strong teams play a game on a neutral field, the probability of a tie will be 0.312, and each side shares a winning probability of 0.344 (put $\beta_i = \beta_j$, K=0.905, and H=1 in formula (2)). This is in accordance with patterns observed in the past. It also indicates that the tie will never have the highest probability of the 3 possible outcomes in any match. The home advantage H is 2.080, indicating on average, a team playing a home game can increase its strength to 2.080 times it's normal strength. The 52 estimated team strength parameters vary between 0 (Andorra and San Marino) and 113.069 (Spain). Spain has been dominating all its European opponents. San Marino and Andorra are the bottom two teams that are even far away from Malta. It's actually not a surprise because neither team has had a victory or a tie during this 4-year period.

If we compare our ranking with the Coca-Cola FIFA ranking published on November 23^{rd}, 2011 (see columns 2 and 3 of Table 1), there are some interesting findings. Eight teams appear in the top 10 of both rankings and there are 19 common teams in both top 20 lists. Similarly, Malta, Andorra, and San Marino are ranked as the 3 bottom teams in both rankings. Even though there are some large differences for some teams, our initial trial seems reasonable. In Section 5.3 we will discuss the similarity among the different rankings in more detail.

4.2. Weighted maximum likelihood estimation

In section 4.1 all the games were treated equally. We differentiated neither between relatively old and new games, nor between officially competitive games and friendly games. In order to make a more realistic ranking, at least two sets of weights should be taken into account: the time effect weight and the type effect weight respectively.

We prefer to treat the time depreciation in a continuous instead of a discrete way as is done in the current FIFA ranking, which uses weights 1 for games played during the past 12 months, 0.5 for games played between the past 24 and 12 months, 0.3 for games played between the past 36 and 24 months, and 0.2 for games played between the past 48 and 36 months. In order to make our ranking comparable to the FIFA ranking, we decide to fix the maximal weight 1 for the games played on the latest match day (November 15^{th}, 2011), as well as the minimal weight 0.2 for the games played on the earliest match day (February 2^{nd},

2008) in our data set. The continuous depreciation function we will use for game i played x_i days going back in time from November 15^{th}, 2011 is $w_{time,i}\,(x_i) = \exp\left(-x_i \log(5)/1382\right)$.

Besides, different tournaments are generally diverse in quality or importance. Especially, the FIFA World Cup final round is regarded the most important soccer tournament all over the world. The continental championships are the most important competitions within each continent. The qualifiers of the two mentioned tournaments are regarded more important than friendly games. Although there are clearly quality differences among different game types, how to evaluate the relative importance is still a somewhat subjective matter. We will employ the weights used in the FIFA ranking, and compare the results with those using equal weights for all games.

When both the time weights ($w_{time,i}$) and type weights ($w_{type,i}$) are included, the weighted likelihood becomes as follows:

$$WL = \prod_{i=1}^{N}\prod_{j=1}^{3}[P_{ij}^{y_{ij}}]^{w_{time,i}w_{type,i}}. \tag{4}$$

Column 3 of Table A1 in appendix presents the results from the WML estimation with the time effect only . The tie effect (0.903) remains almost the same as in the unweighted ML approach (0.905), and the home effect (2.189) is larger now than previously (2.080). England exceeds Netherlands in this ranking, showing an upward trend during the past 4 years. Since its failure in EURO 2008 qualifiers, England had a very good record in both World Cup 2010 qualifiers and friendly games against other European opponents, including some considerable victories against Spain, Germany, Croatia, Denmark, and Sweden. Although Netherlands had a stable performance during the same time interval, with all types of matches equally weighted, their good performance in older competitive matches, especially in EURO 2008 and World Cup 2010, was less influential than in more recent games.

Column 4 of Table A1 in appendix gives the parameter estimates from the WML method with game type effect only. Both the tie effect 0.898 and home effect 2.063 are slightly smaller than that in the unweighted ML approach (0.905 and 2.080 respectively). England dropped 2 positions due to failure in the UEFA EURO 2008 qualifiers and disappointing performance in the FIFA World Cup 2010. Similarly, France also dropped 5 positions due to its poor performance in these two tournaments. In contrast, Spain, Netherlands, and Germany enhanced their leading positions due to stable and convincing performance in these two big events.

Column 6 of Table 1 (or column 5 of Table A1) reports the final WML estimation with both time and type weights. Spain, Netherlands and Germany remain their leading positions. Compared with the type-effects-only results, the home advantage is larger (2.177 v.s. 2.063 in types-only WML), indicating that nowadays teams playing at home can benefit more than before. The tie effect is slightly smaller (0.893 v.s. 0.898), meaning that ties are expected to happen now slightly less often than before. If we compare the final WML results with the time-effects-only approach, the home advantage is slightly smaller (2.177 v.s. 2.189). This indicates that the higher the weights the big events are assigned to, the smaller the home advantage is. This is to be expected as in both the UEFA European Championships or the FIFA World Cup, most games are played on a neutral field (except the host nation(s) involved). The tie effect is also smaller (0.893 v.s. 0.903), indicating that ties happen less common in important tournaments than in friendly games.

5. An evolutional model with stage effects

5.1. Background

When analyzing past matches in the FIFA World Cup and the UEFA European Championships, there seems to be a considerable number of teams which perform very differently in knock-out games and in group stage games. The most famous two teams in this respect are Italy and Germany, the two most successful European nations in the FIFA World Cup. They share the same strategy to not devote all effort to the group stage games. They just want to achieve the minimum target to qualify from their own group as either group winner or runner-up of the 4 teams. In contrast, they will usually play totally devoted at the knock-out stage, trying their best to win the title. The historic data seem to support this view. Italy has won 4 times the FIFA World Cup champion (1934, 1938, 1982, 2006) and 2 times the runner-up (1970, 1994). In the UEFA European Championships, Italy has won once the title (1968), and twice the runner-up (2000, 2012). Italy has a famous tactical style of *concrete defense with breakaway attack*. They benefit from this distinctive strategy especially in the knock-out stage, where every game decides between *life and death*. Before the start of EURO 2012, Italy held the honor that it was never beaten in 90 minutes in any knock-out match either in the UEFA Championships or the FIFA World Cup since the FIFA World Cup 1990! However, in the past 24 years Italy has already lost 5 group stage games: 0-1 to Ireland (1994), 1-2 to Czech Republic (1996), 1-2 to Croatia (2002), 0-3 to Netherlands (2008), and 2-3 to Slovakia (2010) respectively. Italy's *slow start* tradition is somewhat risky. In EURO 1996 and World Cup 2010 they even failed to qualify in their group. Another European giant, Germany, shows a similar pattern as Italy. Germany is the most successful country in the UEFA European Championships history being 3 times champion (1972, 1980, 1996) and 3 times runner-up (1976, 1992, 2008). It's the third most successful nation in the FIFA World Cup (2^{nd} in Europe behind Italy) with 3 times the title (1954, 1974, 1990) and 4 times the runner-up (1966, 1982, 1986, 2002). Since 1954 they have managed to be present in the quarter-final stage (or equivalently the top 8) or further. Similarly to Italy, Germany also prefers to keep their best performance for the knock-out stage games, leading to reservations in their group stage games. In the past 24 years, Germany has lost 6 group stage games in the UEFA EURO Championships or the FIFA World Cup final round: 1-3 to Netherlands (1992), 0-1 to England (2000), 0-3 to Portugal (2000), 1-2 to Czech Republic (2004), 1-2 to Croatia (2008), and 0-1 to Serbia (2010). Germany has been eliminated from the group stage in EURO 2000 and 2004. We regard this divergence between the group and knock-out stage as the effect of their strategy instead of their actual strength, as once they can qualify to the next round, they usually become much more powerful.

5.2. Model specification

In order to capture this strategy, we re-model the 3-way probabilities taking into account that strength grows in later stages of the championships. We assume that teams are always more motivated and devoted than in the previous stage on their way to the final match. To this end, a new variable s_i, the *stage indicator* was created. For UEFA EURO 2008, the group stage games were assigned a value of 1/16, quarter-finals 1/4, semi-finals 1/2, and the final game 1. For FIFA World Cup 2010 there are twice the number of participant teams (32) and therefore one more knock-out stage, the group stage games were assigned a value of 1/16, round of 16 games 1/8, quarter-finals 1/4, semi-finals 1/2, and the final game

Table 2. Kendall tau rank correlation coefficient

τ	FIFA	ML	WML	EVM
FIFA	1.000	0.817	0.876	0.876
ML	-	1.000	0.903	0.903
WML	-	-	1.000	0.997
EVM	-	-	-	1.000

1. For all games other than the final-round of these two tournament, the stage indicator was set equal to 0. We introduce a new parameter G, the strength amplification effect, describing the evolution in performance at different stages. To control the computational complexity, G is the same for all teams. We remodel the 3-way probabilities as follows

$$\begin{aligned} P_{i1} &= \frac{(H\beta_{1i})^{1+s_iG}}{(H\beta_{1i})^{1+s_iG} + {\beta_{2i}}^{1+s_iG} + K\sqrt{(H\beta_{1i}\beta_{2i})^{1+s_iG}}}, \\ P_{i2} &= \frac{K\sqrt{(H\beta_{2i}\beta_{2i})^{1+s_iG}}}{(H\beta_{1i})^{1+s_iG} + {\beta_{2i}}^{1+s_iG} + K\sqrt{(H\beta_{1i}\beta_{2i})^{1+s_iG}}}, \\ P_{i3} &= \frac{{\beta_{2i}}^{1+s_iG}}{(H\beta_{1i})^{1+s_iG} + {\beta_{2i}}^{1+s_iG} + K\sqrt{(H\beta_{1i}\beta_{2i})^{1+s_iG}}}. \end{aligned} \tag{5}$$

We use again weighted maximum likelihood for estimation as in formula (4).

5.3. *Comparisons between different rankings*

Columns 8 and 9 of Table 1 present the results of this evolutional approach (EVM) and the corresponding ranking. The EVM ranking seems quite similar to the WML ranking: the tie effect remains quite close (0.895 v.s. 0.893), and the same holds for the home effect (2.167 v.s. 2.177). We notice that the EVM strengths of teams that participated in neither EURO 2008 nor World Cup 2010 are quite similar to the WML strengths. For teams appearing in at least one of these two tournaments, their strength parameters shrink a little. This is to be expected given that the strength amplification parameter is estimated as 1.430 which indicates considerable strength growth in these championships.

Table 2 reports the pairwise Kendall tau rank correlations among the 4 rankings: FIFA, ML, WML, and EVM. Notice that the WML and EVM rankings are indeed highly correlated (0.997), and they are more in line with the FIFA ranking than the ML ranking. It turns out that both WML and EVM can retrieve the FIFA ranking quite well although the FIFA ranking is based on more than 3000 matches and we only used a subset of 929 matches. Figure 1 and 2 visualize the similarities and differences between the four rankings.

Although EVM is highly correlated with the FIFA ranking, there are also some important differences between the EVM ranking and the FIFA ranking. We discuss and explain the difference for Slovenia, Armenia, Ukraine, and Finland. For Ukraine, the reason of its low position in FIFA ranking (29) is straightforward. As one of the host nations of EURO 2012, Ukraine is exempted from the qualifiers. In other words, since December 2009 (until the start of EURO 2012) all the matches that Ukraine had played were friendly games. The FIFA ranking is points-based, where the friendly games have the lowest weight of 1 in calculating the points. Therefore, compared with the teams which need to play the qualifiers (importance = 2.5), Ukraine has not been able to earn many points in the FIFA ranking.

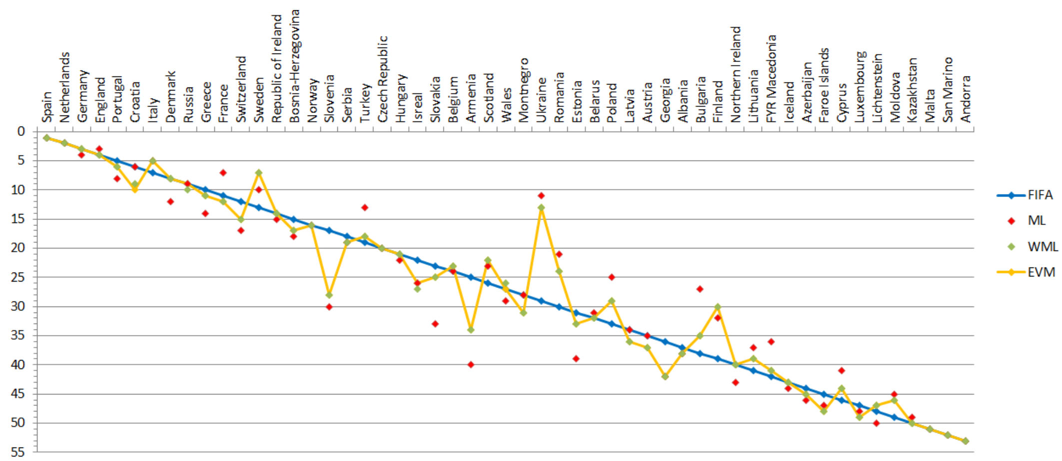

Fig. 1. Comparison of the four rankings

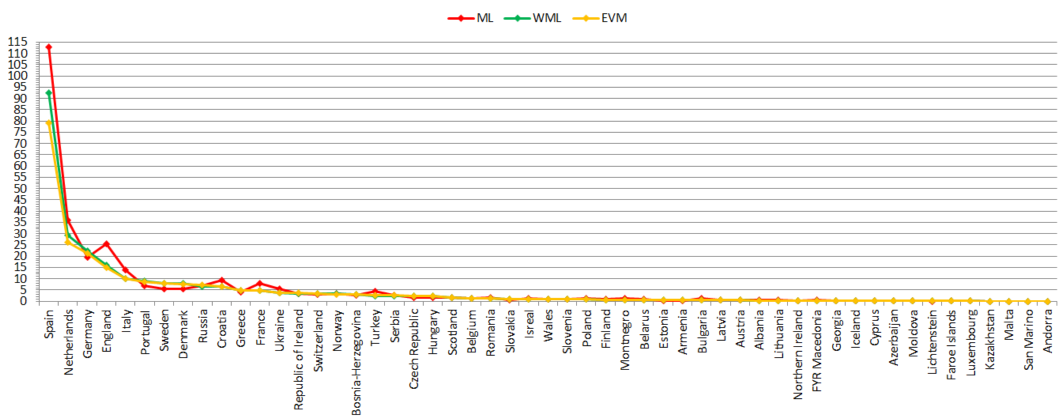

Fig. 2. Comparison of team strength parameters

However, in our model-based ranking, this disadvantage was to some extent relieved. The impact of an individual game on the overall parameter estimation is stronger if a game ends with an unexpected result than otherwise. As Ukraine obtained good results in several friendly games against better teams, their strength improved according to our model. A similar explanation is valid for Poland too.

For Finland, the FIFA ranking has dropped dramatically since October 2010, the start of the the UEFA EURO 2012 qualifiers. Finland lost 5 out of the 6 games against Netherlands, Sweden, and Hungary and ranked only 4^{th} with 10 points for the 10 games in their group. However, with our model-based ranking, a defeat against a stronger opponent is regarded naturally to happen, having less impact on the overall results.

Slovenia benefited in the FIFA ranking from winning some friendly games against non-UEFA teams, including Australia, Algeria, Qatar, New Zealand, which were not included in our subset. As to Armenia, the deviation is harder to explain. On one hand, they played well in the UEFA EURO 2012 qualifiers with 17 points in 10 games (3^{rd} in their group). On the other hand, quite a few of Armenia's victories were against opponents that ranked higher than them, which limited their climbing margin in our model-based ranking.

Generally speaking, the main differences between the FIFA ranking and our EVM ranking can be explained. The two rankings cannot be expected to agree completely as they are not based on exactly the same information. The FIFA ranking uses information of 209 teams' more than 3000 games, whereas we are using only a subset of 929 matches between UEFA international teams. As such the games against teams from other continents in the World Cup 2010 final round are not included in our data set. In the next section we will use the EVM results to analyze the group stage games of EURO 2012 and illustrate its practical use.

6. Analysis of UEFA Euro 2012

As discussed in previous sections, one limitation of the existing FIFA ranking is that it is not effective for making predictions. For a specified game, it's not straightforward to quantify the possibilities of the three possible outcomes based on the FIFA points. Furthermore it cannot take into account the performance difference on home, away or neutral fields. In this section, the estimated parameters of the evolutional model will be used to analyze the UEFA European Football Championship 2012.

6.1. Tournament format

The final tournament was hosted by Poland and Ukraine, between June 8^{th} and July 1^{st} 2012. In the final tournament featured 16 nations. Besides the two host nations, the remaining 14 finalists were determined by the qualifiers. The 16 teams were categorized evenly into 4 pots. For each pot, 4 teams were put in 4 different groups. At the group stage, matches within group A and C were played in Poland, those of group B and D in Ukraine. The 4 Group winners and 4 runner-ups would qualify for the quarter finals. At the knock-out stage, the winner of group A (A1) would face the runner-up of group B (B2), similarly the other 3 quarter-finals would be played between A2-B1, C1-D2, and C2-D1. The first semi-final would be played in Donetsk between the winner of A1-B2 and the winner of C1-D2, while the second semi-final in Warsaw between the winner of A2-B1 and C2-D1. The winners of the two semi-finals would compete for the final title in Kiev.

6.2. Probabilistic analysis

For each UEFA EURO 2012 game, the 3-way outcome (win, tie, and loss) was computed by formula (5). We use the strength values, the home effect, the tie effect and the stage effect estimated by our evolutional model, which takes both time and type weights into account. As there are two host nations Poland and Ukraine, their group stage games are played in their own country, and the home advantage is applied to their 3 group stage games.

Table 3 gives the three-way probabilities for all group stage games of UEFA EURO 2012. These can provide abundant guidance not only for experts, journalists and fans, but also for betting companies to specify the odds. We do not recommend to simply choose the highest probability out of the 3 for a rude prediction, especially when the 3-way probabilities are quite close (e.g., Poland v.s. Czech Republic, Netherlands v.s. Germany, Portugal v.s. Denmark, and Ukraine v.s. Sweden), making it hard to make an effective decision. However, if we make predictions based on the highest probabilities, our model still correctly predicts 13 out of 24 group stage games. In contrast, by always choosing the team with the higher FIFA ranking, 12 correct predictions are made.

7. Conclusion and discussion

In this paper, we developed a new ranking system for soccer. Instead of the points-calculating method used in the current FIFA ranking, we developed a model-based ranking methodology. We modeled the 3-way probabilities in terms of the team strength parameters, a tie effect and a home effect within a multinomial framework. Furthermore, the basic model was improved by taking the time and the game type impact into account. A final improvement was the evolutional model which was designed to distinguish the performance in knock-out matches and group stage games in important tournaments.

The home effect was estimated to be 2.167, expressing the home team benefit. The tie effect was estimated to be 0.895. As this is smaller than 1, ties are expected to occur with a probability smaller than 1/3, which is in line with historic statistics. The stage effect was estimated to be 1.430, indicating a better performance at further stages in a tournament. Although, based on the Kendall tau rank correlation criterion, both the WML and EVM rankings are highly correlated with the existing FIFA ranking, our model-based rankings overcame some limitation of the FIFA ranking by taking into account the home effect, using a more realistic depreciation rate and by providing a tool for making more sound predictions.

There are still some points that should be emphasized. Firstly, as the tie effect is smaller than 1 it is impossible to assign the highest probability to a tie, meaning that we could never predict a tie simply by choosing the highest predicted probability (in accordance with the odds given by betting companies). Secondly, the comparison of our results with the FIFA ranking is not completely valid as our data set only involved the 929 matches played between UEFA teams whereas the FIFA ranking is based on all games played by the 209 FIFA teams. Particularly we lack all the European teams' games against opponents from other associations in the World Cup 2010 final round. Furthermore, our data set contains only games until the end of 2011. We didn't update our data to include the friendly games before the start of the UEFA EURO 2012. As such, the latest information on the 16 teams was missing. Thirdly, no mathematical model can accurately predict internal conflicts in teams during a tournament which may have a large impact too. Of course, a charming feature of football is its randomness. There are various stochastic factors influencing the

Table 3. Analysis of UEFA EURO 2012

Group	Team1	P(T1 wins)	P(Tie)	P(T2 wins)	Team2	FIFA	Actual result
A	Poland	0.177	0.278	**0.545**	Greece	Greece	Tie
A	Poland	0.132	0.255	**0.613**	Russia	Russia	Tie
A	Poland	**0.300**	**0.307**	**0.393**	Czech Rep.	Czech Rep.	Czech Rep.
A	Greece	0.277	**0.305**	**0.418**	Russia	Russia	Greece
A	Greece	**0.498**	0.290	0.212	Czech Rep.	Greece	Czech Rep.
A	Russia	**0.569**	0.270	0.161	Czech Rep.	Russia	Russia
B	Netherlands	**0.579**	0.267	0.154	Denmark	Netherlands	Denmark
B	Netherlands	**0.385**	**0.308**	**0.307**	Germany	Netherlands	Germany
B	Netherlands	**0.557**	0.274	0.169	Portugal	Netherlands	Portugal
B	Germany	**0.517**	0.286	0.197	Portugal	Germany	Germany
B	Germany	**0.540**	0.279	0.181	Denmark	Germany	Germany
B	Portugal	**0.368**	**0.309**	**0.323**	Denmark	Portugal	Portugal
C	Spain	**0.718**	0.207	0.075	Italy	Spain	Tie
C	Spain	**0.833**	0.138	0.029	Ireland	Spain	Spain
C	Spain	**0.769**	0.178	0.052	Croatia	Spain	Spain
C	Italy	**0.422**	**0.304**	0.274	Croatia	Croatia	Tie
C	Italy	**0.542**	0.279	0.179	Ireland	Italy	Italy
C	Ireland	0.237	0.297	**0.466**	Croatia	Croatia	Croatia
D	Ukraine	**0.353**	**0.309**	**0.338**	Sweden	Sweden	Ukraine
D	Ukraine	**0.450**	**0.300**	0.250	France	France	France
D	Ukraine	0.239	0.298	**0.463**	England	England	England
D	Sweden	0.233	0.296	**0.471**	England	England	England
D	Sweden	**0.442**	**0.301**	0.257	France	France	Sweden
D	France	0.163	0.271	**0.566**	England	England	Tie

performance of a team in a certain game, and fortune will always play an important role.

There is of course still room for further improvement: the weights that were used in the weighted maximum likelihood estimation are rather subjective. Looking for sound arguments for their choice or performing a robustness study of the results with respect to the weight allocation might be useful. Also the inclusion of team specific stage effect parameters, expressing that different teams use different strategies, may yield interesting insights.

Appendix

Table A1. All model-based rankings introduced in the paper.

Parameter	ML	WML (time effect)	WML (type effect)	WML (time+type)	EVM
Spain	113.043	47.087	195.881	92.595	79.294
Netherlands	36.144	18.451	53.716	29.253	26.069
Germany	19.686	19.221	26.332	22.209	21.124
England	25.557	20.095	25.464	15.999	14.950
Portugal	6.977	8.351	9.033	9.034	8.721
Croatia	9.415	6.758	11.272	6.693	6.660
Italy	13.943	11.575	13.669	10.123	9.915
Denmark	5.330	5.528	8.599	8.007	7.740
Russia	6.768	4.435	11.076	6.441	7.126
Greece	4.104	3.563	5.162	4.909	4.896
France	7.953	6.335	5.883	4.802	4.763
Switzerland	3.083	2.739	4.032	3.343	3.412
Sweden	5.647	6.155	7.796	8.107	7.831
Republic of Ireland	3.425	3.182	4.232	3.484	3.581
Bosnia-Herzegovina	2.732	2.647	3.555	2.984	2.972
Norway	3.289	2.924	3.686	3.220	3.176
Slovenia	0.866	0.634	1.485	0.924	0.947
Serbia	2.555	1.718	3.810	2.403	2.537
Turkey	4.406	2.859	4.347	2.460	2.553
Czech Republic	1.765	1.923	2.122	2.224	2.230
Hungary	1.560	1.461	2.593	2.219	2.170
Israel	1.234	0.784	1.549	0.998	1.001
Slovakia	0.654	0.504	1.628	1.015	1.093
Belgium	1.415	1.456	1.416	1.202	1.213
Armenia	0.319	0.280	0.508	0.534	0.552
Scotland	1.522	1.895	1.547	1.682	1.659
Wales	1.000	1.000	1.000	1.000	1.000
Montenegro	1.143	0.618	1.228	0.693	0.693
Ukraine	5.605	3.262	6.639	3.810	3.770
Romania	1.649	1.232	1.522	1.116	1.134
Estonia	0.376	0.363	0.694	0.648	0.660
Belarus	0.860	0.663	0.940	0.660	0.660
Poland	1.414	1.139	1.034	0.795	0.804
Latvia	0.616	0.330	0.843	0.469	0.470
Austria	0.542	0.376	0.660	0.453	0.458
Georgia	0.256	0.226	0.279	0.209	0.209
Albania	0.427	0.264	0.466	0.268	0.269
Bulgaria	1.146	0.463	1.215	0.493	0.494
Finland	0.817	0.694	1.057	0.720	0.711
Northern Ireland	0.250	0.134	0.465	0.234	0.240
Lithuania	0.461	0.285	0.483	0.250	0.252
FYR Macedonia	0.465	0.267	0.405	0.225	0.231
Iceland	0.235	0.156	0.286	0.180	0.179
Azerbaijan	0.149	0.127	0.139	0.112	0.113
Faroe Islands	0.069	0.049	0.091	0.064	0.065
Cyprus	0.290	0.129	0.329	0.136	0.135
Luxembourg	0.063	0.050	0.083	0.060	0.060
Liechtenstein	0.046	0.066	0.062	0.076	0.076
Moldova	0.155	0.088	0.135	0.080	0.080
Kazakhstan	0.049	0.050	0.043	0.043	0.044
Malta	0.043	0.029	0.035	0.021	0.021
San Marino	0.000	0.000	0.000	0.000	0.000
Andorra	0.000	0.000	0.000	0.000	0.000
Home effect (H)	2.080	2.189	2.063	2.177	2.167
Tie effect (K)	0.905	0.903	0.898	0.893	0.895
Stage effect (G)					1.430

References

Agresti, A. (1992) Analysis of Ordinal Paired Comparison Data. *Appl. Statist.*, **41**, 287-297.

Bassett, G. W. (1997) Robust Sports Ratings Based on Least Absolute Errors.,*J. Am. Statist. Ass.*, **51**, 99-105.

Bradley, R. A. (1953) Some Statistical Methods in Taste Testing and Quality Evaluation., *Biometrics*, **9**, 22-38.

Bradley, R. A. and Terry, M.E. (1952) Rank Analysis of Incomplete Block Designs I: The Method of Paired Comparisons., *Biometrika*, **32**, 213-232.

David, H.D. (1960) *The Method of Paired Comparisons.*, Newyork: Hafner.

Davidson, R. R. (1970) On Extending the Bradley-Terry Model to Accommodate Ties in Paired Comparison Experiments., *J. Am. Statist. Ass.*, **65**, 317-328.

Dittrich, R., Hatzinger, R., and Katzenbeisser, W. (1998) Modelling the effect of subject-specific covariates in paired comparison studies with an application to university rankings., *Jo. Royal Statist. So. Series C*, **47**, 511-525.

Dyte, D. and Clarke, S.R. (2000) A Rating Based Poisson Model for World Cup Soccer Simulation., *J. Oper. Res. Soc.*, **51**, 993-998.

Fahrmeir, L. and Tutz, G. (1994) Dynamic stochastic models for time-dependent ordered paired comparison systems., *J. Am. Statist. Ass.*, **89**, 1438-1449.

FIFA (2011) The FIFA World Ranking. *http://www.fifa.com/worldranking/index.html*.

Firth, D. (2005) *Bradley-Terry Models in R.*, University of Warwick, United Kingdom.

Glenn, W. A. and David, H. A. (1960) Ties in Paired-Comparison Experiments Using a Modified Thurstone-Mosteller Model., *Biometrics*, **16**, 86-109.

Harville, D. (2003) The Selection or Seeding of College Basketball or football Teams for Postseason Competition., *J. Am. Statist. Ass.*, **89**, 1438-1449.

Kuk A.Y.C. (1995) Modelling paired comparison data with large numbers of draws and large variability of draw percentages among players., *Jo. Royal Statist. So. Series D*, **44**, 523-528.

Lehmann, E. L. (1953) The power of Rank Tests., *Annals of Mathematical Statistics*, **24**, 23-43.

Luce, R. D.(1959) *Individual choice behavior.*, New York: John Wiley & Sons, Inc., Chapter 1.

Mease, D. (2003) A Penalized Maximum Likelihood Approach for the Ranking of College Football Teams Independent of Victory Margin., *Am. Statistn*, **57**, 241-248.

Rao, P. V. and Kuppera, L. L. (1967) Ties in Paired-Comparison Experiments: A Generalization of the Bradley-Terry Model., *J. Am. Statist. Ass.*, **62**, 194-204.

Stefani, R.T. (1997) Survey of the major world sports rating systems., *J. Appl. Stat.*, **24**, 635-646.

Stern, H. S . (1992) Who's Number One? Rating Football Teams, In *Proceedings of the Section on Statistics in Sports*, 1-6.

Stern, S. E. (2011) Moderated Paired Comparisons: a Generalized Bradley-Terry Model for Continuous Data Using a Discontinuous Penalized Likelihood Function, *Appl. Statist.*, **60**, 397-415.

FACULTY OF ECONOMICS AND BUSINESS
Naamsestraat 69 bus 3500
3000 LEUVEN, BELGIË
tel. + 32 16 32 66 12
fax + 32 16 32 67 91
info@econ.kuleuven.be
www.econ.kuleuven.be