Forecasting sports tournaments by ratings of (prob)abilities: A comparison for the EURO 2008

Christoph Leitner*, Achim Zeileis*, Kurt Hornik*

* Department of Statistics and Mathematics, WU Wirtschaftsuniversität Wien, Augasse 2–6, 1090 Wien, Austria

Abstract

Different methods for assessing the abilities of participants in a sports tournament and their corresponding winning probabilities for the tournament, respectively, are embedded into a common framework and compared with respect to their predictive performance. First, ratings of abilities (such as the Elo rating) are complemented by a simulation approach yielding winning probabilities for the full tournament. Second, tournament winning probabilities are extracted from bookmakers odds using a consensus model and the underlying abilities of the competitors are derived by "inverse" application of the tournament simulation. Both techniques are employed for forecasting the results of the *European football championship 2008* (UEFA EURO 2008) for which the consensus model based on bookmakers odds outperforms methods based on the Elo rating and the FIFA/Coca Cola World rating. Moreover, the bookmaker consensus model correctly predicts that the final would be played by teams Germany and Spain (with a probability of about 20.5%) while showing that both finalists profit from being drawn in groups with relatively weak competitors.

Keywords: Sports forecasting, EURO 2008, bookmakers odds, Elo rating, abilities.

1 Introduction

In the course of growing popularity of online sports betting, the analysis and forecasting of competitive sports has been receiving increasing interest. Forecasts of sports events are often based on one of two types of information: ratings or rankings of the competitors' ability/strength, and bookmakers odds for winning a competition of two or more contestants. Here, we show how both types of forecasts—winning probabilities and underlying abilities—can be derived from both sources of information—ability ratings and bookmakers odds. Their predictive performance is assessed in an empirical study forecasting the winner of the European football championship 2008 (UEFA EURO 2008).

Sports ratings or rankings are typically derived by suitably aggregating the competitors' previous performances and are often found to provide predictive power in forecasting tasks. Boulier and Stekler (1999) show that rankings provide forcasting information for basketball tournaments and tennis matches. Lebovic and Sigelman (2001) analyze the predictive accuracy of college football rankings. Suzuki and Ohmori (2008) use the FIFA/Coca Cola World rating (Fédération Internationale de Football Association, 2008), one of the most popular rating system in soccer, as a forecasting tool for the last four FIFA World Cups (1994, 1998, 2002, 2006). In addition, Dyte

This is a preprint of an article published in *International Journal of Forecasting*, **26**(3), 471-481.

doi:10.1016/j.ijforecast.2009.10.001

and Clarke (2000) use the FIFA ratings to predict the distribution of scores in international soccer matches. Another popular rating system is the Elo rating system, originally developed to calculate the relative skills of chess players (e.g., Elo, 2008), which has subsequently also been applied to various other sports including soccer. Song *et al.* (2009) apply it as one method to forecast the winner of single American Football games. Edmans *et al.* (2007) select important soccer games based on the World Football Elo Ratings.

Bookmakers odds represent a rather different type of rating compared to the methods above. Based on the bookmakers' expert judgments (which typically include, but are not limited to, knowledge about past performances) the odds reflect expected outcomes in a particular competition where the bookmakers have strong economic incentives to rate the competitors correctly. A bias (in either direction, too good or too bad) will cost them money, or, in other words, will reduce their profits. Hence, bookmakers can be seen as experts in the matter of sports rating (see Pope and Peel, 1989) and are likely to provide good predictions (Forrest and Simmons, 2000). This is confirmed by various empirical studies in which fixed odds are found to be an efficient forecasting instrument for the outcome of single matches (e.g., Vlastakis *et al.*, 2009; Spann and Skiera, 2009; Song *et al.*, 2007; Forrest *et al.*, 2005b; Dixon and Pope, 2004; Boulier and Stekler, 2003).

One advantage of employing bookmakers odds is that winning probabilities for the corresponding competition can be derived easily while this is not straightforward for many of the ability ratings. However, if abilities are measured on a ratio scale (or can be transformed to such), winning probabilities for pairwise matches can be derived using the approach of the Bradley and Terry (1952) model. Notable in this respect is the Elo rating from which pairwise winning probabilities for single matches can be obtained (e.g., Stefani and Pollard, 2007; Edmans *et al.*, 2007). Thus, when the competition of interest is a single match, forecasts based on ability ratings and bookmakers odds can be compared easily. The same is not true if the competition is a more complex tournament for which the bookmakers odds, by their prospective nature, can include additional effects such as group draws or seedings. To link forecasts of abilities (associated with pairwise winning probabilities) and winning probabilities for sports tournaments, we suggest a simulation approach that allows to (approximately) map abilities to winning probabilities and vice versa.

To compare forecasts based on ability ratings and bookmakers odds, we apply them to the UEFA EURO 2008, one of the world's biggest sports events that took place in June 2008 in Austria and Switzerland. For the odds-based forecasts, quoted bookmakers odds for 16 participating teams were obtained from 45 international bookmakers prior to the tournament (on 2008-04-21) and aggregated in a consensus model. This is compared to the forecasts from the World Football Elo rating (also considered in a note from UBS Wealth Management Research Switzerland, 2008, for prediction of the EURO 2008) and the ranking implied by the FIFA/Coca Cola World rating (also employed in a note from Raiffeisen Zentralbank, 2008), both also obtained on 2008-04-21. Forecasts based on these approaches were first obtained by an analysis in May 2008, prior to the tournament (see Leitner *et al.*, 2008, for a technical report with these preliminary findings), and then reassessed after the end of the tournament. In this ex post comparison, the bookmaker consensus model performs best and predicts the correct final (Germany vs. Spain, with a probability of about 20.5%). Furthermore, the results provide many further insights into the effects of the group draw in the tournament, clearly showing that the two finalists come from groups with relatively weak competitors.

The remainder of this paper is organized as follows: Section 2 discusses some basic features of sports ratings, bookmakers odds, and sports tournaments. Section 3 provides a data and tournament description for the EURO 2008 for which the various forecasts are obtained and assessed in Section 4. Section 5 concludes the paper.

2 Ratings of (prob)abilities in sports tournaments

2.1 Sports ratings

Ratings of "abilities" or "strengths". In competitive sports, players or teams as well as their supporters are interested in ratings of the competitors as a measure of their abilities or strengths. A common strategy for deriving suitable ratings employs adaptive schemes which update assessments based on historic performances upon availability of data about current performances. Typical examples for this include the FIFA/Coca Cola World rating in soccer or the ATP (Association of Tennis Professionals) rating in tennis (see Stefani, 1997, for an overview). Some ratings are based on a simple point system while others employ statistical models, e.g., the Elo rating (Elo, 2008) implies pairwise winning expectancies (see Joe, 1991). A natural application of ability ratings is to employ them for forecasting performances in future matches (e.g., Song *et al.*, 2009). In some sports, ratings are also used for deriving seedings which in turn can be used for forecasting as in Boulier and Stekler (1999).

Bookmakers odds as ratings of winning probabilities. A rather different source of "ratings" of competitors in sports are bookmakers odds: Unlike the ratings discussed above these are not derived directly from past performances but emerge from "expert" knowledge. Of course, this typically encompasses knowledge about past results but may also take into account expectations about future events. Due to the increasing popularity of online sports betting, bookmakers odds are a type of data that is abundant and easily available and that has been successfully employed in forecasts of single matches (e.g., Vlastakis *et al.*, 2009; Spann and Skiera, 2009; Song *et al.*, 2007; Forrest *et al.*, 2005b; Dixon and Pope, 2004; Boulier and Stekler, 2003). Another important difference between bookmakers odds and the ability ratings discussed above is that they are an assessment of outcome probabilities (e.g., winning probabilities in the case of sports tournaments) rather than of the underlying abilities. However, the raw quoted bookmakers odds are no "honest" odds but are the payout amounts for successful bets which has two important implications: (1) They still contain the stake, i.e., the payment for placing the bet (the "1" in Equation 1 below). (2) More importantly, the bookmakers odds contain a profit margin, the so-called "overround", which means that the "true" underlying odds are actually larger (see e.g., Henery, 1999; Forrest *et al.*, 2005b). Assuming that the overround δ is constant across all possible outcomes (e.g., the same for all competitors winning a tournament), it can be computed by restricting the corresponding probabilities to sum to unity. More precisely, the raw quoted odds $rawodds_i$ for event i can be adjusted to $odds_i$ and then transformed to probabilities p_i via:

$$odds_i = (rawodds_i - 1)\,\delta, \qquad (1)$$

$$p_i = 1 - \frac{odds_i}{1 + odds_i}. \qquad (2)$$

Then, δ can be chosen such that $\sum_i p_i = 1$. (Note, that the complementary probabilities have to be used as the bookmakers odds represent expectations for an outcome not to occur.) In the case of winning odds for a tournament, this means that the implied winning probabilities can be easily derived from the quoted odds for all competitors.

2.2 Sports tournaments

Pairwise comparisons. In many sports disciplines, winners and losers are determined by pairwise comparisons, called matches or games. Clearly, the outcome from a match depends on the current abilities of the two competitors. Given abilities measured on a ratio scale, the classical method for computing winning probabilities from abilities is the Bradley and Terry (1952) approach which derives the probability for competitor i beating competitor j as:

$$\pi_{i,j} = \frac{ability_i}{ability_i + ability_j} \qquad (i \neq j), \tag{3}$$

where $ability_i$ is the ability for team i on a ratio scale. However, for many sports rating systems it is not clear what the underlying measurement scale is. A notable exception is the Elo rating (Elo, 2008) which uses a similar approach for obtaining winning expectancies. Hence, as discussed in detail below, Elo ratings can easily be transformed to abilities in the sense of Equation 3.

Tournament schedule. If a winner shall be determined from a group (rather than just a pair) of competitors, this is typically accomplished by using a sequence of pairwise comparisons, called tournament. Various designs are available for constructing suitable schedules for such a tournament (see Scarf and Bilbao, 2006, for a discussion). In a round-robin tournament, where each competitor (or player or team) plays each other, it is obvious that the strongest competitor has the highest winning probability in each pairwise comparison and therefore the highest chance to win the tournament, followed by the second strongest competitor and so on. However, for other tournament schedules the strongest competitor does not necessarily have the highest probability of winning. For example, if the tournament schedule is based on a draw of a group phase and/or a knockout phase, some competitors might be favored/discriminated by being drawn together with relatively weak/strong competitors. However, when the tournament schedule and the abilities of its participants are known, it is (in principle) straightforward to compute the associated winning probabilities based on the pairwise probabilities from Equation 3 by applying conditional probabilities to all possible tournament "paths". As explicit enumeration of all paths can be burdensome, the winning probabilities can also be approximated easily by simulating a large number of tournament runs (100,000 say) and then assessing the empirical winning proportions $\widetilde{p}$ for each competitor:

$ability$ abilities of all competitors	$\rightarrow$	$\pi_{i,j}$ pairwise winning probabilities for all matches	$\rightarrow$	100,000 runs tournament simulations	$\rightarrow$	$\widetilde{p}$ simulated winning probabilities for tournament

The resulting (approximated) winning probabilities $\widetilde{p}(ability)$ then also capture all "tournament effects" induced by the schedule. Note that this approach models the contestants' abilities as constant over the course of the competition and might be further enhanced to accommodate hypothesized patterns of change in abilities. Also, this generic simulation setup might require adaptation to some details of a specific tournament, e.g., for EURO 2008 potential ties after the group phase need to be resolved (as described in detail in Section 4).

3 EURO 2008: Data and tournament description

3.1 Data

Elo ratings. The World Football Elo Ratings (Advanced Satellite Consulting Ltd, 2008), Elo ratings for short, for all 16 teams participating in the EURO 2008 have been collected from `http://www.eloratings.net/` (accessed 2008-04-21). In contrast to many other sports rating systems (such as the FIFA ratings below), the Elo ratings imply winning expectancies for pairwise comparisons (see Elo, 2008, Equation 46). The probability that team i beats team j can be related to

$$\pi_{i,j} = \frac{1}{10^{-(Elo_i - Elo_j)/400} + 1} \qquad (i \neq j), \tag{4}$$

where Elo_i and Elo_j are the Elo ratings for teams i and j, respectively. For home teams (i.e., Austria and Switzerland in the EURO 2008), 100 rating points are added to the Elo rating (Advanced Satellite Consulting Ltd, 2008). Thus, the Elo ratings are essentially on a $\log_{10}$ scale which is somewhat different from the standard Bradley and Terry (1952) model. However, using Equations 3 and 4, it is easy to provide a transformation to log-abilities in the Bradley-Terry sense which imply the same pairwise winning probabilities $\pi_{i,j}$. As the log-abilities are just defined up to a constant γ, we choose γ such that they are on a logit scale:

$$\log\left(ability_i^{(ELO)}\right) = \frac{\log(10)}{400} Elo_i + \gamma, \tag{5}$$

$$\sum_i \text{logit}^{-1}\left(\log\left(ability_i^{(ELO)}\right)\right) = 1, \tag{6}$$

where Equation 6 implies $\gamma = -13.496$ for the EURO 2008 data, log is the natural logarithm, and logit^{-1} denotes the inverse of the logit function. The resulting Elo log-abilities are provided in Table 1 where the logit scale facilitates comparison with logits of tournament winning probabilities derived in the following.

Bookmakers odds. Longterm odds for winning the EURO 2008 were obtained from the websites of 45 international bookmakers for all 16 participating teams on 2008-04-21. These are all of 50 European top-selling online sports bookmakers who offered odds for this event. Prior to all further analysis, the odds are adjusted by removing the stake and a bookmaker-specific overround (see Equation 1) and then transformed to winning probabilities by means of Equation 2. This yields tournament winning probabilities $p_{i,b}$ for $i = 1, \ldots, 16$ teams and $b = 1, \ldots, 45$ bookmakers which reflect the bookmakers' beliefs about the outcome of the EURO 2008.

FIFA ratings. The FIFA/Coca Cola World ratings (Fédération Internationale de Football Association, 2008), FIFA ratings for short, for all 16 participating teams were retrieved from `http://www.fifa.com/` on 2008-04-21. These ratings capture abilities of the teams but on an unknown scale so that it is not straightforward to compute pairwise winning probabilities $\pi_{i,j}$ or tournament winning probabilities p_i (see McHale and Davies, 2007, for an approach for building more complex statistical models based on the FIFA rating). Therefore, in the following, the FIFA ratings are employed only for comparison as a ranking (rather than rating).

3.2 Tournament

The UEFA EURO 2008 is a tournament where 52 European national football teams (UEFA's members) compete in a multi-stage modus (qualification, group and knockout stage) to determine the European champion. First, 16 teams are determined via a qualification stage for the group stage, i.e., the main EURO 2008 tournament carried out in June 2008 in Austria and Switzerland. Table 1 lists the 16 teams as drawn into four groups, labeled A through D. Each group of four plays a round-robin—every team plays every other team, for a total of six matches within the group—and the top two teams in each group advance to the next stage, the quarter-final. The winner of group A plays against the second best team of group B (first quarter-final) and the winner of group B plays against the second best team of group A (second quarter-final). Analogously, the winner of group C plays against the second best team of group D (third quarter-final) and the winner of group D plays against the second best team of group C (forth quarter-final). The four winners of the quarter-finals reach the semi-finals, where the winner of the first quarter-final plays against the winner of the second one and the winner of the third quarter-final plays against the winner of the forth. The winners of the semi-finals then play the final and the winner of the final is the European football champion (Union of European Football Associations, 2009).

4 Forecasting of the EURO 2008

In this section, forecasts of team (log-)abilities and winning probabilities for the EURO 2008 tournament are obtained based on the Elo ratings and the bookmakers odds, respectively. The resulting four quantities are compared with the actual result of the tournament and the best-performing method is analyzed in some more detail.

4.1 Forecasting based on the Elo ratings

As argued in Sections 2 and 3, the Elo ratings Elo_i ($i = 1, \dots, 16$) represent an assessment of the current ability/strength of the teams participating the EURO 2008. By construction, pairwise probabilities $\pi_{i,j}$ for all combinations of participants can be obtained. Furthermore, to approximate winning probabilities that include tournament effects such as the group draw, the empirical winning proportions from 100,000 simulated tournaments are used:

$$ability_i^{(ELO)} = \exp\left(\frac{\log(10)}{400} Elo_i - 13.496\right), \tag{7}$$

$$p_i^{(ELO)} = \widetilde{p}\left(ability^{(ELO)}\right)_i. \tag{8}$$

Thus, $ability^{(ELO)}$ is the vector of abilities (in the Bradley-Terry sense) based on which the tournament simulations are carried out. The results for all teams are reported in Table 1.

By adopting the classical Bradley-Terry model, the simulation of each match yields only a winner and a loser without the possibility of a tie and without further information about the number of goals or the goal difference. This is sufficient for the knock-out stage of the tournament as it reflects that the actual matches always have a winner (if necessary in overtime and penalties). However, for the group phase within the simulation this approach might result in tied teams. If necessary, we resolve such ties by additional "fictitious" matches between the tied teams to obtain unique winners and the runner-ups of the groups.

Our simulation method could be extended by using more elaborate simulation techniques including ties and number of goals, e.g., a model where the team scores follow independent Poisson distributions (e.g., Maher, 1982; Dixon and Coles, 1997; Dyte and Clarke, 2000), or an ordered probit regression model (Goddard and Asimakopoulos, 2004).[1]

According to the Elo rating, Italy is the strongest team ($\log(ability^{(ELO)}) = -1.97$) and also has the highest probability for winning the tournament ($p^{(ELO)} = 18.28\%$). However, the second strongest team France has only the third highest winning probability ($\log(ability^{(ELO)}) = -2.09$, $p^{(ELO)} = 14.08\%$) while Germany is only the fifth strongest but has the second highest winning probability ($\log(ability^{(ELO)}) = -2.34$, $p^{(ELO)} = 15.99\%$). Thus, team Germany clearly profits from being drawn in a group (B) with weaker competitors while France has a certain disadvantage from being placed in a group (C) with strong competitors such as Italy. This tournament effect can be conveniently assessed by comparing differences between the teams' log-abilities and their winning logits, respectively (as both measurements have been constructed such that they are on a logit scale). For example, Italy's margin over Germany of 0.37 ($= -1.97 - (-2.34)$) is reduced to 0.16 ($= -1.5 - (-1.66)$) by including tournament effects while France's margin over Germany of 0.25 is reversed to -0.15. Furthermore, it is worth noting that team Spain, the favorite in group D, has the fourth highest winning probability ($p^{(ELO)} = 13.14\%$) while Austria has the lowest chances of winning the EURO 2008 ($p^{(ELO)} = 0.14\%$), nonwithstanding its potential home advantage (see e.g., Forrest *et al.*, 2005a; Clarke and Norman, 1995).

[1] However, all approaches should give reasonable approximations of the probabilities for being promoted to the next round.

	$\log(ability_i)$		$p_i(\%)$		$\text{logit}(p_i)$		
	ELO	BCM	ELO	BCM	ELO	BCM	Group
Germany	−2.34	−2.33	15.99	17.45	−1.66	−1.55	B
Spain	−2.25	−2.41	13.14	12.21	−1.89	−1.97	D
Italy	−1.97	−2.40	18.28	11.34	−1.50	−2.06	C
Portugal	−2.95	−2.54	3.36	9.97	−3.36	−2.20	A
France	−2.09	−2.50	14.08	9.14	−1.81	−2.30	C
Netherlands	−2.33	−2.62	8.29	6.77	−2.40	−2.62	C
Croatia	−2.86	−2.77	5.03	6.72	−2.94	−2.63	B
Czech Republic	−2.67	−2.74	7.17	5.88	−2.56	−2.77	A
Switzerland	−2.79	−2.88	5.18	3.92	−2.91	−3.20	A
Greece	−2.93	−2.91	2.76	3.31	−3.56	−3.37	D
Sweden	−3.32	−2.98	0.77	2.87	−4.86	−3.52	D
Russia	−3.42	−3.00	0.55	2.72	−5.20	−3.58	D
Turkey	−3.27	−3.06	1.30	2.26	−4.33	−3.77	A
Romania	−2.72	−3.04	2.77	2.12	−3.56	−3.83	C
Poland	−3.35	−3.19	1.19	2.05	−4.42	−3.87	B
Austria	−3.93	−3.85	0.14	0.93	−6.55	−4.67	B

Table 1: Log-abilities, winning proabilities, and corresponding logits of all teams for the EURO 2008 based on the Elo rating (ELO) and on the bookmaker consensus model (BCM). The ELO log-abilities are directly computed from the Elo ratings and winning probabilities are derived via simulation. The BCM logits are estimated by team-wise means of bookmaker log-odds, the corresponding log-abilities are found by "inverse" simulation. The rows are sorted by the BCM winning probabilities.

4.2 Forecasting based on bookmakers odds

When appropriately adjusted and transformed, as described in Sections 2 and 3, the bookmakers odds yield expected winning probabilities $p_{i,b}$ for each team $i = 1, \dots, 16$ and bookmaker $b = 1, \dots, 45$. In the following, a single forecast for the winning probability of each team is obtained by aggregation of the $p_{i,b}$ across bookmakers. Subsequently, a vector of underlying team abilities is found by "inverse" application of the simulation approach adopted above.

The bookmakers odds are prospective ratings of the performance of the 16 participating teams in the EURO 2008 which vary between 45 bookmakers. To obtain an aggregated measure for each team some sort of consensus between the different ratings has to be computed (for discussions of various strategies for the aggregation of forecasts see e.g., Song *et al.*, 2009, 2007; Kolb and Stekler, 1996; Schnader and Stekler, 1991; Zarnowitz and Lambros, 1987). Here, we adopt a simple additive model on a logit scale

$$\text{logit}(p_{i,b}) \quad = \quad \text{logit}(p_i) + \epsilon_{i,b}, \tag{9}$$

where p_i is the latent winning probability for team i and $\epsilon_{i,b}$ is the deviation of bookmaker b for team i. In principle, it is possible to refine this model further by including group effects into the winning logits $\text{logit}(p_i)$ or bookmaker-specific bias and variance into the deviation $\epsilon_{i,b}$. See Leitner *et al.* (2009) for an exploration of several mixed-effects models (e.g., Pinheiro and Bates, 2000) capturing different team- and bookmaker-specific effects. However, as the bookmakers' expectations about the EURO 2008 are rather homogeneous a straightforward fixed-effects model with zero-mean deviations $\epsilon_{i,b}$ should be appropriate. Thus, the consensus winning logits are simply means across bookmakers:

$$\widehat{\mathrm{logit}(p_i)} \quad = \quad \frac{1}{45} \sum_{b=1}^{45} \mathrm{logit}\,(p_{i,b})\,. \tag{10}$$

Transforming these logits back to the probability scale yields the bookmakers' consensus winning probabilities $p_i^{(BCM)}$. Both probabilities and corresponding logits for this bookmakers consensus model (BCM), are shown in Table 1. The model captures 98.21% of the variance of the $p_{i,b}$, the associated estimated standard error of $\epsilon_{i,b}$ is 0.11396.

Although forecasting the winning probabilities for the EURO 2008 is the main concern in our investigation, there is also interest in the team abilities underlying the bookmakers' expectations. The tournament schedule was already known at the time the bookmakers odds were retrieved, and hence should be included in the expectations about the outcome of the tournament. To strip the "tournament effects" (see Section 2.2) from this measure, we employ an "inverse" application of the simulation approached described in the previous sections. More precisely, we want to find a set of team abilities $\mathit{ability}_i$ ($i = 1, \dots, 16$) that result in simulated winning probabilities $\widetilde{p}(\mathit{ability})_i$ that are as similar as possible to the consensus winning probabilities $p_i^{(BCM)}$:

$$p_i^{(BCM)} \quad = \quad \mathrm{logit}^{-1}\left(\widehat{\mathrm{logit}(p_i)}\right), \tag{11}$$

$$\mathit{ability}^{(BCM)} \quad = \quad \underset{\mathit{ability}}{\mathrm{argmin}} \sum_{i=1}^{16} \left| p_i^{(BCM)} - \widetilde{p}\,(\mathit{ability})_i \right|. \tag{12}$$

The minimum in the second line is determined using a local search strategy for the full vector $\mathit{ability}^{(BCM)}$ where 100,000 tournament runs are employed in each evaluation of $\widetilde{p}(\cdot)$. The results are reported in Table 1.

According to the BCM, Germany has the highest chances of winning the EURO 2008 ($p^{(BCM)} =$ 17.45%) with some margin over Spain (12.21%) and Italy (11.34%). Thus, although there is considerable overlap among the top five teams obtained from BCM and Elo results, the ranking and associated winning probabilities of these teams are rather different. Also, France (which was the second strongest team according to the Elo rating) has only the fifth largest winning probability (9.14%). Finally, host country Austria is again expected to have the lowest winning probability (0.93%) but it is somewhat larger in absolute terms compared to the Elo forecast.

To investigate the tournament effect, differences in the teams' winning logits can again be compared with differences in their log-abilities. Again, this shows that Germany greatly profits from the group draw because its margin in terms of winning logits over Spain or Italy (0.42 and 0.51, respectively) is greatly reduced in terms of log-abilities (0.08 and 0.07). Note also that this reduction is larger for Italy than for Spain, conveying that Italy suffers particularly from being drawn in the strong group C (often referred to as the "group of death").

4.3 Ex post comparison of all forecasts

The previous subsections present two different types of forecasts (abilities and winning probabilities) derived from two different types of ratings (Elo rating and bookmakers odds). As usual in forecasting, it is of central interest which strategy performs best in practice. Although this is difficult to answer because there are no "real" replications of the tournament, we can compare the forecasts with the single real outcome of the EURO 2008.

Table 2 assesses the predictive performance of all four forecasts by comparing them with the actual tournament outcomes using Spearman's rank correlation. For the actual results, a total ranking including ties is employed, as commonly used in rankings of such incomplete tournaments.[2]

[2]Various strategies for dissolving the ties have been explored but did not lead to qualitatively different results.

	$p^{(BCM)}$	$ability^{(BCM)}$	$p^{(ELO)}$	$ability^{(ELO)}$	FIFA rating
Tournament ranking	0.525	0.441	0.304	0.203	0.373
$p^{(BCM)}$		0.988	0.871	0.771	0.809
$ability^{(BCM)}$			0.909	0.826	0.841
$p^{(ELO)}$				0.956	0.809
$ability^{(ELO)}$					0.815

Table 2: Spearman's rank correlation between the actual tournament ranking and rankings according to the estimated BCM winning probabilities and (log-)abilities, simulated Elo winning probabilities and (log-)abilities (equivalent to the original Elo rating), and the FIFA/Coca Cola World rating.

First, this shows that the winning probabilities (including the tournament effects) have higher correlation with the actual outcome (0.525 for BCM and 0.304 for ELO, respectively) compared to the corresponding (log-)abilities (0.441 and 0.203). Second, the forecasts based on the bookmakers odds clearly outperform those based on the Elo ratings. This conveys that the prospective ratings of experts (i.e., the bookmakers) have been more useful than the retrospective performance-based Elo ratings.

In addition to the four forecasts derived in this paper, Table 2 also provides correlations with the ranking implied by the FIFA/Coca Cola World rating. Interestingly, this has a higher Spearman correlation (0.373) with the tournament outcome than the Elo forecasts. Furthermore, it is more closely associated with both (log-)ability measurements (0.841 and 0.815) than with the corresponding winning probabilities (0.809 and 0.809). This confirms that the (retrospective) FIFA rating is an assessment of the teams' current ability and conveys that its predictive power could be enhanced if the corresponding winning probabilities could be computed or simulated. However, as no rigorous method for computing pairwise winning probabilities $\pi_{i,j}$ based on the FIFA rating is known to us, we cannot pursue this approach here.

To investigate the sources of the good performance of the BCM for the winning probabilities, it is useful to extract the two best-ranked teams from each group in Table 1. This shows that the consensus winning probabilities correctly predict five teams (Germany, Spain, Italy, Portugal, Croatia) which played the quarter-finals, as well as the actual final (played by the teams Germany and Spain). The big surprises of the tournament were teams Russia and Turkey which both reached the semi-finals rather unexpectedly. Whereas the BCM ranked team Russia better than the Elo and the FIFA rating, the converse is true for team Turkey. Furthermore, France surprisingly did not reach the quarter-finals which was neither expected by the bookmakers nor using the Elo or FIFA ratings. However, it was somewhat more likely using the BCM.

4.4 Tournament analysis based on the BCM forecast

In addition to the team abilities and winning probabilities (Table 1), some further insights can be gained from the best-performing BCM forecast due to adoption of the simulation approach. So far, we have only considered the empirical winning proportions of each team in the 100,000 tournament runs. But, of course, the empirical proportions of reaching the quarter-final, semi-final, and final can be extracted as well. Figure 1 shows the performance of each team in the simulations based on $ability^{(BCM)}$ as a performance curve (or "survival" curve over the course of the tournament). The endpoints of the curves are the simulated winning probabilities, which are by construction (Equation 12) (roughly) identical to the probabilities derived from the BCM (Table 1).

The performance curves in Figure 1 show that groups B and D are rather heterogeneous with weaker teams and clear favorites (Germany and Spain, respectively) while groups A and C are

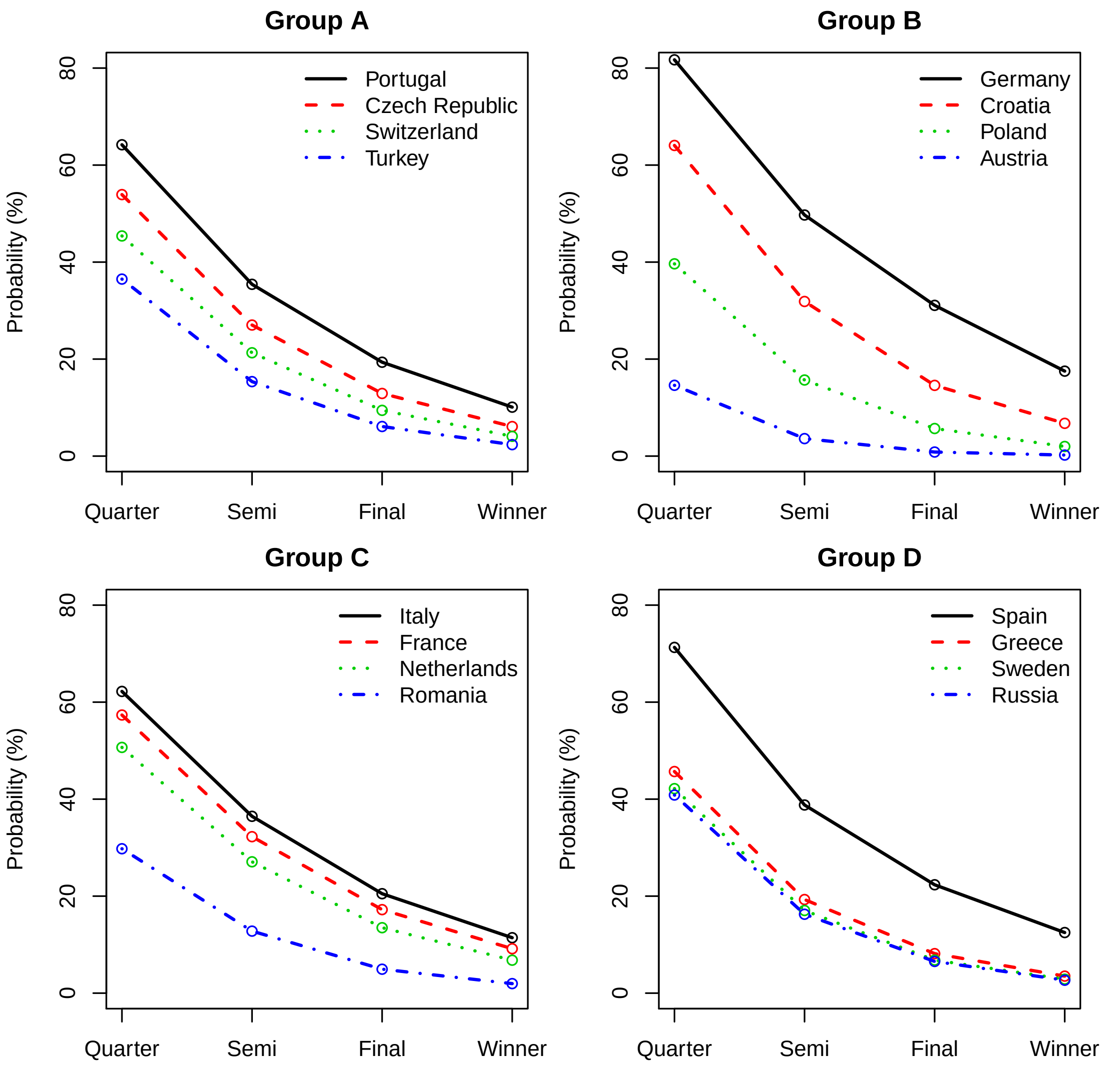

Figure 1: Simulated probabilities (from 100,000 tournament runs based on the BCM consensus abilities) for reaching the quarter-final, the semi-final, the final, and for winning the EURO 2008.

rather homogeneous. This group effect can also be quantified on an aggregated level by considering deviations of the mean group winning logits (computed from Table 1) from the overall mean winning logits across all teams. Despite the fact that group B includes the bookmakers' favorite of winning the European championship (Germany), group B clearly is the weakest group and has the smallest chance to include the winner (with a deviation of -0.187 on the logit scale). This is followed by group D with a deviation of -0.116. Group C, on the other hand, is clearly the toughest group and has the greatest probability of including the champion (0.293). Group A can be interpreted as the average group with a deviation of 0.010 from the overall mean.

The simulation also provides information about the most likely coupling for the final: A match of Germany and Spain, the actual final, occurs with the highest probability of 20.45%. Given this coupling in the final, the winning probabilities of both teams are given by the Bradley-Terry model (Equation 3) based on the teams' estimated abilities $ability_i^{(BCM)}$. Although team Germany has a slight advantage with a winning probability of 52.08%, this essentially conveys that no clear favorite exists in this final. This is confirmed by the actual EURO 2008 final which ended with a very close result: Germany 0, Spain 1.

5 Summary and outlook

We embedded various methods for rating players/teams in competitive sports into a common framework that allows for forecasting winning probabilities in sports tournaments (rather than single matches) and obtaining the competitors' underlying strengths/abilities. The link between abilities and winning probabilities is established by means of a simulation approach that takes into account potential tournament effects such as group draws or seedings. Specifically, these methods are applied to the World Football Elo rating and the odds from a set of international bookmakers and assessed using forecasts of the European football championship 2008. A consensus model for the bookmakers odds performs best in this comparison, correctly predicting the actual final of the tournament and revealing clear tournament effects due to the group draw.

Although the model forecasts provide promising results for the EURO 2008, various improvements are conceivable and deserve further study: The tournament simulation could be enhanced to provide not only winners and losers but more realistic results (such as goals or goal differences in a soccer tournament). The bookmaker consensus model adopted here only includes a fixed team effect but could be extended to encompass further fixed or random effects capturing, for example, group strengths, bookmaker bias, or differences in variance.

Furthermore, our results convey that the prospective rating based on aggregated expert judgment in the bookmaker consensus model provides more accurate forecasts of sports tournament outcomes compared to retrospective ratings that derive current team/player abilities from past performances. Application of both approaches to future tournaments will continue to explore the potential of these methods and help to establish a more complete picture.

Computational details

All computations were carried out in the R system (version 2.10.1) for statistical computing (R Development Core Team, 2009).

References

Advanced Satellite Consulting Ltd (2008). "The World Football Elo Rating System." [Online; accessed 2008-04-21], URL http://www.eloratings.net/.

Boulier BL, Stekler HO (1999). "Are Sports Seedings Good Predictors?: An Evaluation." *International Journal of Forecasting*, **15**, 83–91.

Boulier BL, Stekler HO (2003). "Predicting the Outcomes of National Football League Games." *International Journal of Forecasting*, **19**, 257–270.

Bradley RA, Terry ME (1952). "Rank Analysis of Incomplete Block Designs: I. The Method of Paired Comparisons." *Biometrika*, **39**, 324–345.

Clarke SR, Norman JM (1995). "Focus on Sport – Home Ground Advantage of Individual Clubs in English Soccer." *Journal of the Royal Statistical Society D*, **44**(4), 509–521.

Dixon M, Coles S (1997). "Modelling Association Football Scores and Inefficiencies in the UK Football Betting Market." *Journal of the Royal Statistical Society C*, **46**(2), 265–280.

Dixon M, Pope P (2004). "The Value of Statistical Forecasts in the UK Association Football Betting Market." *International Journal of Forecasting*, **20**, 697–711.

Dyte D, Clarke S (2000). "A Rating Based Poisson Model for World Cup Soccer Simulation." *The Journal of the Operational Research Society*, **51**(8), 993–998.

Edmans A, García D, Norli O (2007). "Sports Sentiment and Stock Returns." *Journal of Finance*, **62**(4), 1967–1998.

Elo AE (2008). *The Rating of Chess Players, Past and Present*. Ishi Press, San Rafael, United States.

Fédération Internationale de Football Association (2008). "FIFA/Coca-Cola World Ranking." [Online; accessed 2008-04-21], URL http://www.fifa.com/.

Forrest D, Beaumont J, Goddard J, Simmons R (2005a). "Home Advantage and the Deabate about Competitive Balance in Professional Sports Leagues." *Journal of Sports Sciences*, **23**(4), 439–445.

Forrest D, Goddard J, Simmons R (2005b). "Odds-Setters as Forecasters: The Case of English Football." *International Journal of Forecasting*, **21**, 551–564.

Forrest D, Simmons R (2000). "Forecasting Sport: The Behaviour and Performance of Football Tipsters." *International Journal of Forecasting*, **16**, 317–331.

Goddard J, Asimakopoulos I (2004). "Modelling Football Match Results and the Efficiency of Fixed-Odds Betting." *Journal of Forecasting*, **23**, 51–66.

Henery RJ (1999). "Measures of Over-Round in Performance Index Betting." *Journal of the Royal Statistical Society D*, **48**(3), 435–439.

Joe H (1991). "Rating Systems Based on Paird Comparison Models." *Statistics & Probability Letters*, **11**, 343–347.

Kolb R, Stekler HO (1996). "Is There a Consensus among Financial Forecasters." *International Journal of Forecasting*, **12**, 455–464.

Lebovic JH, Sigelman L (2001). "The Forecasting Accuarcy and Determinants of Football Rankings." *International Journal of Forecasting*, **17**, 105–120.

Leitner C, Zeileis A, Hornik K (2008). "Who is Going to Win the EURO 2008? (A Statistical Investigation of Bookmakers Odds)." *Report 65*, Department of Statistics and Mathematics, Wirtschaftsuniversität Wien, Research Report Series. URL http://epub.wu.ac.at/.

Leitner C, Zeileis A, Hornik K (2009). "Forecasting the Winner of the UEFA Champions League 2008/09." In R Koning, P Scarf (eds.), *Proceedings of the 2nd International Conference on Mathematics in Sport – IMA Sport 2009*, pp. 94–99. ISBN: 979-0-905091-21-1.

Maher M (1982). "Modelling Association Football Scores." *Statistica Neerlandica*, **36**, 109–118.

McHale I, Davies S (2007). "Statistical Analysis of the Effectiveness of the FIFA World Rankings." In J Albert, RH Koning (eds.), *Statistical Thinking in Sports*, pp. 77–90. Chapman & Hall/CRC, Boca Raton, Florida.

Pinheiro J, Bates D (2000). *Mixed-Effects Models in S and S-PLUS*. Statistics and Comupting. Springer-Verlag New York, New York, USA. ISBN 0-387-98957-9.

Pope PF, Peel DA (1989). "Information, Prices and Efficiency in Fixed-Odds Betting Market." *Economica*, **56**, 323–341.

R Development Core Team (2009). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org/.

Raiffeisen Zentralbank (2008). "RZB-Analyse: Wer wird Fußball-Europameister?" [Online; accessed 2008-05-09], URL http://www.rzb.at/.

Scarf P, Bilbao M (2006). "The Optimal Design of Sporting Contests." *Report 320*, Salford Business School, Working Paper Series. URL http://www.mams.salford.ac.uk/.

Schnader M, Stekler HO (1991). "Do Consensus Forecasts Exist?" *International Journal of Forecasting*, **7**, 165–170.

Song C, Boulier BL, Stekler HO (2007). "The Comparative Accuracy of Judgmental and Model Forecasts of American Football Games." *International Journal of Forecasting*, **23**, 405–413.

Song C, Boulier BL, Stekler HO (2009). "Measuring Consensus in Binary Forecasts: NFL Game Predictions." *International Journal of Forecasting*, **25**, 182–191.

Spann M, Skiera B (2009). "Sports Forecasting: A Comparison of the Forecast Accuracy of Prediciton Markets, Betting Odds and Tipsters." *Journal of Forecasting*, **28**, 55–72.

Stefani RT (1997). "Survey of the Major World Sports Rating Systems." *Journal of Applied Statistics*, **24**(6), 635–646.

Stefani RT, Pollard R (2007). "Football Rating Systems for Top-Level Competition: A Critical Survey." *Journal of Quantitative Analysis in Sports*, **3**(3), 1–20.

Suzuki K, Ohmori K (2008). "Effectiveness of FIFA/Coca-Cola World Ranking in Predicting the Results of FIFA World Cup Finals." *Football Science*, **5**, 18–25.

UBS Wealth Management Research Switzerland (2008). "European Champions for 2008 Will Be. . . ." [Online; accessed 2008-04-21], URL http://www.ubs.com/.

Union of European Football Associations (2009). "UEFA Champions League." [Online; accessed 2009-04-23], URL http://en.euro2008.uefa.com/tournament/.

Vlastakis N, Dotsis G, Markellos RN (2009). "How Efficent is the European Football Betting Market? Evidence from Arbitrage and Trading Strategies." *Journal of Forecasting*, **28**, 426–444. doi:10.1002/for.1085.

Zarnowitz V, Lambros LA (1987). "Consensus and Uncertainty in Economic Prediction." *Journal of Political Economy*, **95**, 561–621.

A Bookmakers odds

	AT	HR	CZ	FR	DE	GR	IT	NL
bwin	81.00	14.00	15.00	10.00	4.75	26.00	7.00	14.00
X888	81.00	13.00	13.00	8.50	5.00	26.00	8.00	12.00
bet365	101.00	12.00	13.00	9.00	4.50	23.00	8.00	13.00
betdirect	101.00	13.00	15.00	9.00	5.00	21.00	7.00	11.00
bet1128	91.00	14.00	17.00	8.50	5.50	25.00	7.50	12.00
betChronicle	104.00	14.00	14.00	9.20	5.00	29.00	7.90	13.50
betfred	101.00	13.00	13.00	9.00	5.00	26.00	8.00	13.00
betinternet	101.00	13.00	15.00	9.50	5.00	26.00	8.00	13.00
better	101.00	13.00	15.00	9.00	5.00	21.00	7.00	11.00
bluesq	81.00	13.00	13.00	8.50	5.00	26.00	8.00	12.00
boylesports	81.00	13.00	15.00	10.00	5.00	26.00	7.50	15.00
canbet	101.00	13.00	15.00	9.00	4.75	26.00	6.50	12.00
centrebet	101.00	11.00	15.00	10.00	5.00	26.00	8.00	15.00
coral	81.00	13.00	15.00	8.00	4.50	26.00	7.00	13.00
ladbrokes	81.00	13.00	15.00	9.00	5.00	26.00	7.00	13.00
lasseters	101.00	13.00	15.00	9.00	5.00	26.00	7.00	13.00
paddypower	67.00	13.00	13.00	9.00	5.00	26.00	8.00	13.00
pagebet	101.00	12.00	13.00	9.00	5.00	26.00	8.00	13.00
partybets	81.00	11.00	11.00	10.00	5.50	21.00	7.00	10.00
skybet	67.00	13.00	17.00	9.00	5.00	26.00	8.00	11.00
sportingbet	101.00	13.00	15.00	10.00	4.50	21.00	7.00	11.00
stanjames	101.00	13.00	15.00	9.00	5.00	21.00	7.00	11.00
totesport	101.00	13.00	15.00	9.00	4.50	26.00	7.50	13.00
vcbet	81.00	11.00	13.00	9.00	4.50	26.00	8.00	13.00
hill	81.00	11.00	12.00	10.00	5.00	26.00	8.00	13.00
pinalesports	83.82	12.24	15.33	10.13	5.10	29.88	7.37	12.35
expekt	67.00	12.00	13.00	9.00	5.00	26.00	7.50	10.00
gamebookers	81.00	11.00	11.00	10.00	5.00	21.00	7.00	10.00
betathome	90.00	12.00	15.00	10.00	5.00	25.00	8.00	14.00
gera	81.00	12.00	13.00	10.00	4.75	23.00	7.00	13.00
sunmarker	100.00	13.00	15.00	9.00	5.00	25.00	7.50	13.00
noxwin	100.00	13.00	15.00	9.00	5.00	25.00	7.50	13.00
betway	81.00	11.00	13.00	10.00	5.00	23.00	8.00	13.00
betsafe	100.00	12.00	13.00	10.00	5.00	26.00	8.00	13.00
betboo	81.00	13.00	15.00	8.00	4.25	21.00	6.50	11.00
intertops	101.00	12.00	15.00	9.00	4.00	26.00	8.00	14.00
unibet	100.00	12.50	15.00	10.00	5.00	33.00	8.00	12.50
mybet	81.00	13.00	13.00	9.00	5.00	26.00	7.50	13.00
betsson	100.00	12.00	15.00	8.00	5.00	30.00	8.00	10.00
nordicbet	75.00	12.00	15.00	10.00	5.40	25.00	5.50	12.00
digibet	100.00	13.00	15.00	9.00	5.00	27.00	7.50	13.50
betclick	80.00	14.00	16.00	10.00	5.00	25.00	8.00	14.00
admiralbet	100.00	12.00	15.00	10.00	5.00	20.00	8.00	12.00
interwetten	100.00	12.00	16.00	7.50	5.00	30.00	7.50	12.00
bet24	100.00	12.50	15.00	10.00	5.35	30.00	7.50	12.50

Table 3: Published longterm odds of 45 international bookmakers for the teams of group A (Czech Republic CZ, Portugal PT, Switzerland CH, Turkey TR) and of group B (Austria AT, Croatia HR, Germany DE, Poland PL) for winning the EURO 2008 (source: websites of the bookmakers, online, accessed 2008-04-21).

	PL	PT	RO	RU	ES	SE	CH	TR
bwin	41.00	8.50	34.00	29.00	7.00	29.00	23.00	41.00
X888	29.00	9.00	41.00	29.00	7.50	26.00	15.00	41.00
bet365	41.00	8.50	41.00	34.00	7.00	29.00	21.00	34.00
betdirect	51.00	8.00	41.00	21.00	7.00	26.00	17.00	34.00
bet1128	42.00	9.00	41.00	35.00	7.50	28.00	23.00	37.00
betChronicle	43.00	8.30	39.00	41.00	7.10	31.00	27.00	50.00
betfred	34.00	8.00	51.00	34.00	7.00	34.00	21.00	41.00
betinternet	41.00	8.50	41.00	34.00	7.00	29.00	21.00	41.00
better	51.00	8.00	41.00	21.00	7.00	26.00	17.00	34.00
bluesq	29.00	9.00	41.00	29.00	7.50	26.00	15.00	41.00
boylesports	41.00	8.00	29.00	26.00	7.00	34.00	21.00	34.00
canbet	46.00	8.00	26.00	31.00	6.00	34.00	26.00	41.00
centrebet	29.00	8.50	51.00	34.00	7.20	34.00	23.00	41.00
coral	51.00	9.00	29.00	34.00	7.00	29.00	23.00	34.00
ladbrokes	34.00	9.00	41.00	29.00	7.00	26.00	21.00	34.00
lasseters	41.00	8.50	34.00	31.00	7.00	31.00	21.00	41.00
paddypower	34.00	8.50	41.00	29.00	7.00	26.00	19.00	34.00
pagebet	51.00	8.50	41.00	29.00	7.00	29.00	23.00	41.00
partybets	34.00	8.50	29.00	26.00	7.00	29.00	19.00	29.00
skybet	41.00	8.00	51.00	26.00	7.00	26.00	21.00	34.00
sportingbet	51.00	8.50	34.00	34.00	6.50	34.00	21.00	41.00
stanjames	51.00	8.00	41.00	21.00	7.00	26.00	17.00	34.00
totesport	41.00	9.00	51.00	34.00	7.00	26.00	21.00	34.00
vcbet	41.00	8.50	34.00	26.00	6.50	26.00	17.00	34.00
hill	41.00	8.00	41.00	26.00	7.00	26.00	21.00	34.00
pinalesports	41.40	8.97	41.40	36.11	7.17	35.57	23.46	41.40
expekt	29.00	9.00	34.00	29.00	7.00	26.00	21.00	29.00
gamebookers	34.00	8.50	29.00	26.00	7.00	29.00	19.00	29.00
betathome	40.00	8.50	40.00	30.00	7.00	30.00	22.00	34.00
gera	39.00	8.50	34.00	29.00	7.00	29.00	21.00	34.00
sunmarker	45.00	8.00	40.00	45.00	6.00	30.00	25.00	45.00
noxwin	45.00	8.00	40.00	45.00	6.00	30.00	25.00	45.00
betway	34.00	8.50	34.00	26.00	7.00	26.00	21.00	34.00
betsafe	45.00	8.50	40.00	32.00	7.00	32.00	22.00	40.00
betboo	41.00	8.50	34.00	31.00	6.50	26.00	21.00	34.00
intertops	41.00	8.50	41.00	34.00	6.50	29.00	23.00	51.00
unibet	40.00	8.50	45.00	35.00	7.00	33.00	27.00	45.00
mybet	41.00	8.00	34.00	29.00	6.80	29.00	26.00	41.00
betsson	40.00	9.00	50.00	30.00	8.00	25.00	22.00	35.00
nordicbet	40.00	9.00	40.00	30.00	7.00	25.00	25.00	30.00
digibet	50.00	8.00	50.00	40.00	6.50	33.00	27.00	50.00
betclick	40.00	8.00	40.00	35.00	7.00	30.00	20.00	40.00
admiralbet	40.00	8.00	40.00	30.00	8.00	30.00	20.00	20.00
interwetten	50.00	10.00	50.00	30.00	7.50	30.00	20.00	30.00
bet24	50.00	9.25	50.00	35.00	7.00	35.00	25.00	40.00

Table 4: Published longterm odds of 45 international bookmakers for the teams of group C (France FR, Italy IT, Netherlands NL, Romania RO) and of group D (Greece GR, Russia RU, Spain ES, Sweden SE) for winning the EURO 2008 (source: websites of the bookmakers, online, accessed 2008-04-21).

B Elo and FIFA Ratings

	Elo	FIFA
Italy	2003	1391
France	1981	1205
Spain	1953	1319
Netherlands	1940	1124
Germany	1938	1261
Czech Republic	1881	1234
Romania	1872	1085
Croatia	1847	1041
Greece	1836	1187
Portugal	1832	1125
Turkey	1777	860
Sweden	1768	858
Poland	1762	839
Switzerland	1760	621
Russia	1751	851
Austria	1562	323

Table 5: Published World Football Elo ratings (http://www.eloratings.net/) and FIFA/Coca Cola World ratings (http://www.fifa.com/), accessed 2008-04-21.