

Who is Going to Win the EURO 2008? A Statistical Investigation of Bookmakers Odds

Leitner, Christoph; Zeileis, Achim; Hornik, Kurt

DOI:

[10.57938/47ea6a20-58db-41ed-a7c3-143f61b1f090](https://doi.org/10.57938/47ea6a20-58db-41ed-a7c3-143f61b1f090)

Published: 01/01/2008

Document Version

Publisher's PDF, also known as Version of record

[Link to publication](#)

Citation for published version (APA):

Leitner, C., Zeileis, A., & Hornik, K. (2008). *Who is Going to Win the EURO 2008? A Statistical Investigation of Bookmakers Odds*. Research Report Series / Department of Statistics and Mathematics No. 65
<https://doi.org/10.57938/47ea6a20-58db-41ed-a7c3-143f61b1f090>

Who is Going to Win the EURO 2008? (A Statistical Investigation of Bookmakers Odds)



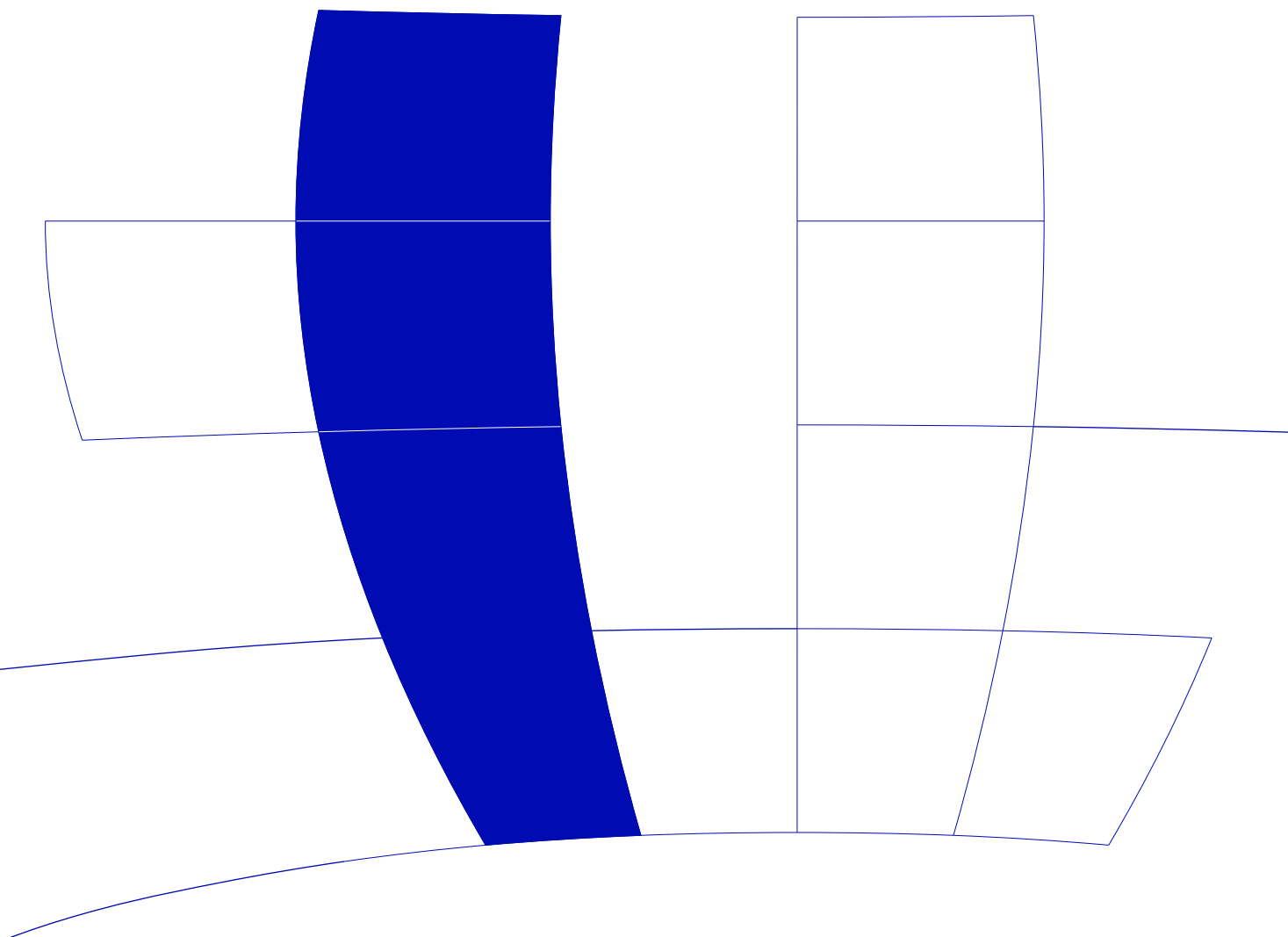
Christoph Leitner, Achim Zeileis, Kurt Hornik

Department of Statistics and Mathematics
Wirtschaftsuniversität Wien

Research Report Series

Report 65
May 2008

<http://statmath.wu-wien.ac.at/>



Who is going to win the EURO 2008?

(A statistical investigation of bookmakers odds)

Christoph Leitner*, Achim Zeileis*, Kurt Hornik*

* Department of Statistics and Mathematics, Wirtschaftsuniversität Wien, Augasse 2–6, A-1090 Vienna, Austria

Abstract

This June one of the biggest and most popular sports tournaments will take place in Austria and Switzerland, the *European soccer championship 2008* (UEFA EURO 2008). Therefore millions of soccer fans in Europe and throughout the world are asking themselves: “Who is going to win the EURO 2008?”

Many people, including sports experts and former players, give their guesses and expectations in the media, but there is also a group with financial incentives, like some economists who expect economical increases for the country of the winning team and bookmakers and their customers who directly make money with their beliefs. Some predictions are only guesses, but other predictions are based on quantitative methods, such as the studies of UBS Wealth Management Research Switzerland and the Raiffeisen Zentralbank. In this report we will introduce a new method for predicting the winner.

Whereas other prediction methods are based on historical data, e.g., the Elo rating, or the FIFA/Coca Cola World rating, our method is based on current expectations, the bookmakers odds for winning the championship. In particular we use the odds for winning the championship for each of the 16 teams of 45 international bookmakers. By interpreting these odds as rating of the expected strength of the teams by the bookmakers, we derive a consensus rating by modelling the log-odds using a random-effects model with a team-specific random effect and a bookmaker-specific fixed effect. The consensus rating of a team can be used as an estimator for the unknown “true” strength of a team.

Our method predicts team Germany with a probability of about 18.7% as the EURO 2008 winner. We predict also that the teams playing the final will be Germany and Spain with a probability of 13.9%, where Germany will win with a probability of 55%. In our study, Italy, the favorite according to the current FIFA/Coca Cola World ranking and Elo ranking, has a much lower probability than these teams to win the tournament: only 10.6%. The defending champion Greece has low chances to win the title again: about 3.4%. Furthermore, the expected performance of the host countries, Austria and Switzerland, is much better in the bookmakers consensus than in the retrospective Elo and FIFA/Coca Cola World ratings, i.e., indicating an (expected) home court advantage. Despite the associated increase in the winning probabilities, both teams have rather poor chances to win the tournament with probabilities of 1.3% and 4.0%, respectively. In a group effect study we investigate how much the classification into the four groups (A–D) affects the chance for a team to win the championship.

Keywords: EURO 2008, odds, latent variable, random-effects model, consensus rating.

1 Introduction

“Football is a simple game; 22 men chase a ball for 90 minutes and at the end, the Germans always win.” Gary Lineker

The main result of our study, predicting team Germany as the winner of EURO 2008, will not come as a surprise for many soccer fans and it could be wrapped up much more expressively as in the (in-)famous quote of Gary Lineker. But is this really the end of the story already? From a statistical point of view, it would clearly be desirable to complement such a result with the corresponding probability, leading to further questions: e.g., who is the expected runner-up, which teams will reach quarter- and semi-finals and with which probabilities. Here, we suggest a new method for answering these questions and compare them with other approaches.

One strategy is to use historical or retrospective data on scores in previous games as in the Elo rating (Advanced Satellite Consulting Ltd, 2008) or the FIFA/Coca Cola World rating (Fédération Internationale de Football Association, 2008). Additionally variables like the overall value of the team players, performance in past tournaments, or performance during the qualification could be included (e.g., Raiffeisen Zentralbank, 2008; UBS Wealth Management Research Switzerland, 2008). In contrast, our method uses prospective data, namely bookmakers odds from 45 international bookmakers for winning the championship for each of the 16 teams. These odds can be interpreted as ratings of the expected strength of the teams by the bookmakers. Using this rating data we derive a consensus rating by modelling the log-odds using a random-effects model. The result can be used as an estimator for the unknown “true” strength of a team. The motivation for using bookmakers odds are twofold: (1) As indicated above, they are an assessment of the expected performance at this specific tournament (rather than the performance in other games). (2) The bookmakers have strong economic incentives to rate the teams correctly because this is how they earn their money. A rating bias (in either direction, too good or too bad) will cost them money, or, in other words, will reduce their profits.

The provided odds of the bookmakers do not represent the true chances that the event will occur, but are the amounts that the bookmaker will pay out on winning bets. In formulating the odds to display the bookmaker will have included a profit margin which effectively means that the payout to a successful punter is less than that represented by the true chance of the event occurring. This profit is known as the “over-round” on the “book”. Profiting in gambling involves predicting the relationship of the true probabilities to the payout odds. The odds or amounts the bookmaker will pay are determined by the amounts bet on each of the respective possible events. They reflect the balance of wagers on either side of the event, and include the deduction of a bookmakers brokerage fee (for further details see Wikipedia, 2008).

For this study we use the published longterm odds for winning the European championship of 45 international bookmakers. We collected the data from all bookmakers’ websites (Online: accessed 2008-04-21). The whole dataset is included in the Appendix A in Table 4 and Table 5.

The paper is organized into four sections: Section 2 gives a description of the method which is applied in Section 3, and Section 4 discusses the used method and the results.

2 Method

Our method assumes a relationship between the bookmakers odds and the “true” strength of a team. To estimate this true strengths, we first adjust the published odds $\widehat{odds}_{i,j}$ of bookmaker j for team i into the underlying beliefs $odds_{i,j}$ by removing the over-round in Section 2.1. From these odds we obtain an estimate of the latent strength \widehat{odds}_i for each team i using a one-factor random-effects model (e.g., Pinheiro and Bates, 2000) in Section 2.2.

2.1 Pre-processing

Because of the fact that the bookmakers do not provide the true chances that a team will win the tournament, we have to adjust the published odds for over-round in order to recover the underlying beliefs of the bookmakers. We assume that the over-round is constant for each bookmaker across all teams, i.e.

$$\widetilde{odds}_{i,j} = \alpha_j odds_{i,j} \quad (1)$$

where $odds_{i,j}$ is the real belief of bookmaker j for team i and α_j is the proportion that bookmaker j pays in case of a win (i.e., the reciprocal value of the over-round).

This odds can be transformed to a probability scale via

$$p_{i,j} = 1 - \frac{odds_{i,j}}{1 + odds_{i,j}} \quad (2)$$

where $p_{i,j}$ is the expected winning probability of bookmaker j for team i . Note, that we have to use complementary probabilities as the bookmakers odds represent expectations for losing the tournament.

For computation of the pay-out proportion α_j for each bookmaker j we use the constraint that the sum of all probabilities has to be one:

$$\sum_i p_{i,j} = 1 \quad \forall j. \quad (3)$$

2.2 Modeling

Using a multi-rater panel of 45 international bookmakers we propose a stochastic model for the rating error and the underlying odds distribution. For this model, the odds representing probabilities have to be transformed into a continuous variable. For further calculations we use therefore the log-odds, which represents the logits of the probabilities of Equation 2.

The relationship between the estimated noisy log-odds $\log(odds)_{i,j}$ and the latent score $\log(odds_i)$ can be written as

$$\log(odds_{i,j}) = \log(odds_i) + \epsilon_{i,j} \quad (4)$$

where $\epsilon_{i,j}$ is the rating error (on the log-odds scale) of bookmaker j for team i .

We assume that the rating error depend due to different rating systems and different levels of information of the bookmaker only on the rater. Therefore, we introduce a one-factor random-effects model where the rating errors are independent from the true score (log-odds).

The relationship between the estimated noisy log-odds $\log(odds)_{i,j}$ and the latent score $\log(odds_i)$ is then

$$\log(odds_{i,j}) = \log(odds_i) + \mu_j + \sigma_j Z_{i,j} \quad (5)$$

where $odds_{i,j}$ is the real belief of rater j for team i , μ_j is the rating bias to the mean score of bookmaker j , σ_j^2 is the variance of the rating error of bookmaker j and Z_{ij} is the standardized rating error of bookmaker j for team i .

Under normality and independence assumptions the parameters of the model 5 can be fitted via maximum likelihood, yielding: $\log(\widehat{odds}_i)$, $\widehat{\mu}_j$, and $\widehat{\sigma}_j$.

The estimated log-odds $\log(\widehat{odds}_i)$ can be interpreted as the consensus rating for team i and can be therefore used to build a consensus ranking of all participating teams of the EURO 2008 or to calculate pairwise winning probabilities to simulate all match results of the tournament.

Group A	Czech Republic (CZ)	Portugal (PT)	Switzerland (CH)	Turkey (TR)
$\log(\widehat{odds}_i)$	2.72	2.20	3.14	3.69
\widehat{odds}_i	15.19	9.06	23.00	40.02
\widehat{p}_i (%)	6.18	9.94	4.17	2.44
Group B	Austria (AT)	Croatia (HR)	Germany (DE)	Poland (PL)
$\log(\widehat{odds}_i)$	4.57	2.59	1.66	3.77
\widehat{odds}_i	96.35	13.33	5.24	43.46
\widehat{p}_i (%)	1.03	6.98	16.03	2.25
Group C	France (FR)	Italy (IT)	Netherlands (NL)	Romania (RO)
$\log(\widehat{odds}_i)$	2.30	2.09	2.61	3.73
\widehat{odds}_i	10.00	8.07	13.59	41.64
\widehat{p}_i (%)	9.09	11.03	6.85	2.35
Group D	Greece (GR)	Russia (RU)	Spain (ES)	Sweden (SE)
$\log(\widehat{odds}_i)$	3.30	3.51	2.00	3.44
\widehat{odds}_i	27.07	33.33	7.41	31.18
\widehat{p}_i (%)	3.56	2.91	11.88	3.11

Table 1: Consensus ratings of all teams in $\log(\widehat{odds}_i)$, with the corresponding odds \widehat{odds}_i and probabilities \widehat{p}_i for all teams.

3 Results

According to our method described above, we compute the consensus ratings (Section 3.1), compare them with the common rating systems, the Elo rating and FIFA/Coca Cola World rating systems (Section 3.2) and use them to generate pairwise winning probabilities for all teams (Section 3.3). With these probabilities we simulate all matches of the tournament and predict the winner of the European championship 2008 (Section 3.4).

In addition to that prediction we provide some interesting results about the group effects (Section 3.5) and the estimated performance of all teams (Section 3.4). Section 3.6 concludes our results with some results about the bookmakers, the raters.

3.1 Team strength: Consensus rating

Table 1 shows the estimated consensus ratings $\log(\widehat{odds}_i)$ and corresponding in odds \widehat{odds}_i and in probabilities \widehat{p}_i which can be calculated according to Equation 2 for win the championship for all participating teams grouped by the origin groups of the first round of the tournament.

According to the consensus ratings of Table 1 team Germany has with average odds of 5.24 for winning the tournament the highest probability to win (16.03%) the top favorite of winning the European championship 2008. With this rating team Germany is the clearly the top favorite of group B. The team of Austria, one of the host countries has with odds of 96.36 and the corresponding probability of 1.03% clearly the lowest chance to win the EURO 2008.

The favorite of group A is not the group member with home advantage, team Switzerland. With a consensus odds of 23.00 team Switzerland has a rather low probability (4.17%) to win the tournament, but the team of the last host country Portugal has a good chance to win the title ($\widehat{odds} = 9.06$, $\widehat{p} = 9.94\%$).

The favorite of group C team Italy has with odds of 7.41 the third largest probability (11.03%) to win the whole tournament, but group member France follows with odds of 10 and a probability of 9.09% closely.

Team	DE	ES	IT	PT	FR	HR	ND	CZ
Bet consensus ranking	1	2	3	4	5	6	7	8
Bet consensus rating	1.000	0.881	0.852	0.812	0.778	0.679	0.673	0.635
Elo ranking	5	3	1	10	2	8	4	6
Elo rating	0.853	0.887	1.000	0.612	0.950	0.646	0.857	0.723
FIFA/Coca Cola World ranking	3	2	1	7	5	10	8	4
FIFA/Coca Cola World rating	0.878	0.933	1.000	0.751	0.826	0.672	0.750	0.853
Team	CH	GR	SE	RU	TR	RO	PL	AT
Bet consensus ranking	9	10	11	12	13	14	15	16
Bet consensus rating	0.492	0.436	0.387	0.365	0.302	0.288	0.273	0
Elo ranking	14	9	12	15	11	7	13	16
Elo rating	0.449	0.621	0.467	0.429	0.488	0.703	0.454	0
FIFA/Coca Cola World ranking	15	6	12	13	11	9	14	16
FIFA/Coca Cola World rating	0.279	0.809	0.501	0.494	0.503	0.713	0.483	0

Table 2: Comparison of the bet consensus ratings $\log(\widehat{odds}_i)$ with the Elo rating and the FIFA/Coca Cola World rating in terms of ranks and scaled ratings. All three ratings are scaled to the unit interval where the rating of the best team is set to one and the rating of the weakest team is set to zero.

The second top favorite of the EURO 2008 is team Spain ($odds = 7.41$, $p = 11.88\%$) which is a member of the first round group D. Another group member of group D, the defending champion team Greece has with odds 27.07 and a probability of 3.56% rather low chances to win the title again.

3.2 Comparison of the consensus rating with common ratings

Using the estimated consensus rating we can rank the teams, like the common FIFA/Coca Cola World ranking according to Fédération Internationale de Football Association (2008) and the Elo ranking system developed by Arpad Elo to rate chess players, adapted by Bob Runyan (1997) for international soccer (Advanced Satellite Consulting Ltd, 2008). Both rankings employ historical data and include the past teams' results of a given period according to a special weighting key including the importance of the game.

Table 2 compares our consensus ranking with the Elo ranking from 2008-04-21 (Advanced Satellite Consulting Ltd, 2008) and the FIFA/Coca Cola World ranking from 2008-04-09 (Fédération Internationale de Football Association, 2008) of all championship participants where the ratings of the three rating systems are scaled to the unit interval. The team with the lowest rating is set to zero, whereas the team with the highest (best) rating is set to one. Therefore the teams can be compared according to their ranks and their relative rating distances.

The comparison of the different rankings shows that the bookmakers' top favorite team Germany is not the best-rated team by the Advanced Satellite Consulting Ltd (2008) and the Fédération Internationale de Football Association (2008). Whereas the FIFA rates team Germany currently on the third place, team Germany is only on the fifth rank of the current Elo ranking. According to both retrospective rankings Italy is currently the best European team. We assume that the bookmakers include the information that Germany's performance at the last world championship two years ago (third place) was better than expected and the German soccer society has proclaimed the mission "Euro champion 2008".

Whereas the rating distance of team Switzerland to favorite Germany is very similar according to the consensus rating and the Elo rating, Switzerland is clearly better ranked according to the bookmakers odds than by the FIFA/Coca Cola World ranking. According to the FIFA/Coca Cola World rating system, team Switzerland is the second weakest team (behind Austria) at the tournament. The main

reason for the better ranking and rating of team Switzerland according to the bookmakers consensus is the home advantage. As a co-host country of the championship team Switzerland plays all their games of the first round in Basel.

Despite the fact that Austria is a host country too, team Austria is according to the consensus ranking, as well as according to the Elo and FIFA/Coca Cola World ranking the rank outsider of the tournament. However, the rating distance to the next team is in the case of the consensus rating closer than in the other two rating systems what signifies a small included home advantage too.

Another team (Russia) is considerably better ranked according to the consensus rating depending on the bookmakers odds than by the Elo ranking and the FIFA/Coca Cola World ranking. We assume that the reason for this difference is the coach effect. Russia has with Guus Hiddink a very popular and experienced coach. He has lead many teams to high performances in tournaments and championships.

On the other side Romania has a much worse consensus rank than in the Elo ranking or FIFA/Coca Cola World ranking. A potential explanation could be the fact that this team is a member of the “group of death”. The bookmakers assume that Romania has the smallest chance to survive this group. We assume that if Romania would be a participant in another group they will have much better odds than in this group C .

3.3 Pairwise winning probabilities

Soccer is a sport as many others where winner and loser are determined by pairwise comparisons, called match or game. In the common soccer leagues every team plays against every other team over the season at least once. For example in the Austrian soccer league the ten league teams play four times against each other (two home games for each team) in a season. At the end of a season (after almost a year) the winner of the league is determined by a ranking scheme which counts the wins (three points), the ties (one point), the defaults and the scores of all games.

In a tournament like the European championships, there is a group phase and play-offs so that not every team play against each other. Thus, the tournament schedule might change the winning probabilities compared to those derived directly from the consensus rating (see Table 1). To include effects stemming from the schedule in the winning probabilities, we suggest to simulate the tournament based on pairwise winning probabilities for all combinations of teams. These can be derived from scores for the team strength such as our consensus rating $\log(\widehat{odds})$. The probability that team A beats team B can be computed as

$$\hat{p}_{A,B} = \frac{\log(\widehat{odds}_B)}{\log(\widehat{odds}_A) + \log(\widehat{odds}_B)} \quad (A \neq B). \quad (6)$$

Note that this approach accounts for wins and losses but excludes ties and also does not yield information about the expected goal difference in a match of A and B .

For example, the probability that team Germany wins against team Austria is 73.39%, in almost 3 of 4 games Germany will be the winner of the game. This seems very high for Germany, but the Austrian fans know from wretched experience how difficult it is to beat the “big brother” at a major tournament. In fact, Austria has succeeded in this only once, 1978, in the famous 3:2 in Cordoba at the world championship in Argentina.

All pairwise winning probabilities according to Equation 6 for each team to win against each other team using the consensus ratings (in log-odds) of Table 1 are shown in Figure 1 which have been placed in five color-categories.

The relationship plot of the pairwise winning probabilities for the 16 teams (Figure 1) figures out that team Germany has a good probability (greater than 55%) to beat all other teams, except Spain. Spain

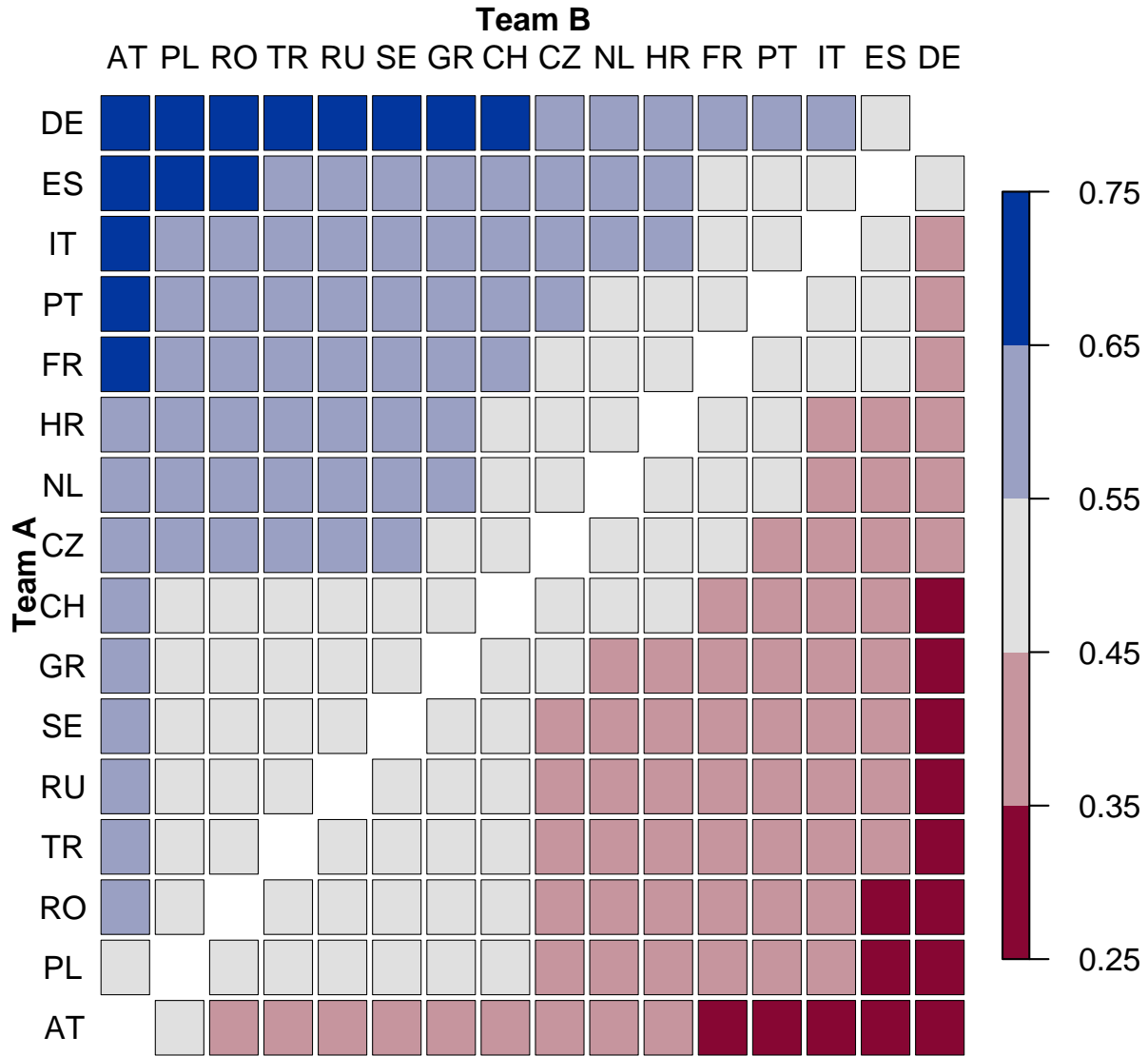


Figure 1: Probability that team A beats team B using the consensus ratings of teams A and B (see Table 1).

and Italy follow with rather high winning probabilities against many other teams. There is a subgroup of six teams (Switzerland, Greece, Sweden, Russia, Turkey, Romania, and Poland) with rather balanced winning chances against each other. The best chance to win a game for team Austria is with almost a equal chance to win (45.23%) or lose the competition with team Poland.

3.4 Probability to win the European championship

As pointed out above we use the pairwise winning probabilities of Section 3.3 to simulate all games of the tournament. There are mainly two motivations for this: (1) to obtain probabilities not only for

Ranking	1	2	3	4	5	6	7	8
Team	DE	ES	IT	PT	FR	HR	ND	CZ
Winning probability	18.69	12.23	10.56	9.80	8.23	6.79	6.05	5.93
Ranking	9	10	11	12	13	14	15	16
Team	CH	GR	RU	SE	PL	TR	RO	AT
Winning probability	3.99	3.42	3.01	3.00	2.41	2.38	2.20	1.31

Table 3: Simulated probability (100,000 simulations) to win the European championship 2008 for each of the 16 teams by simulating the tournament matches based on the pairwise winning probabilities (see also Figure 1).

winning the tournament but also for reaching the quarter-final, semi-final and final (this section); (2) to recover the group effects (introduced by the tournament schedule) from the consensus rating $\log(\widehat{odds}_i)$ (Section 3.5).

The 16 participating teams of the EURO 2008 are divided into four groups. In the first round of the tournament each team plays against each other team of a group. This leads to three games for each team. The best two teams of each group are promoted to the second round, the quarter-final. The winner of group A plays against the second best team of group B (first quarter final) and reversely (second quarter-final). For the promoted teams of group C and D the similar modus is used (third and fourth 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 quarter-final and the winner of the third quarter-final plays against the winner of the fourth quarter-final. The winner of the semi-finals play than the final. The final winner is than called the winner of the European championship 2008.

In our simulation scenarios there are only a winner and a loser of a game, there is no tie. If there are ties in number of winnings in group at the end of the first round, we simulate one or more “fictitious” games between the tie teams to solve this problem. In the playoffs (quarter-finals and semi-finals) and the final there are no ties and we can therefore easily simulate the game results (the winners and loser) according to our method.

The simulation of the whole tournament is done 100,000 times and the probability to win the European championship for each of the 16 teams is computed (Table 3).

The simulation result of all tournament games predicts Germany with a winning probability of 18.69% as the winner of the title. Germany’s winning chance is therefore clearly better than of the strongest competitor Spain, which has a winning probability of 12.23%. A comparison of these simulation results with the transformed probabilities of the consensus log-odds \widehat{p}_i of Table 1 shows that the simulated probabilities of Germany and Spain are higher than the transformed probabilities. This difference could be explained by a impact of the group classification (four groups), so-called group effect which is explained in Section 3.5.

Italy, the currently best European team according to the FIFA/Coca Cola World rating and the Elo rating system has with a tournament winning probability of 10.56% a much lower chance to win than team Germany, the top team according to our method. The defending champion team Greece has with a probability of 3.42% a rather low chance to win the title again. Despite the home court advantage of both host country teams Austria and Switzerland, both teams are with winning probabilities of 1.31% and 3.99% are clear outsiders to win the championship.

In addition to predict the chances for all teams to win the title, we can also compute probabilities for some tournament events, e.g., the most frequent final, or the performance of all teams. Therefore, we predict that the final playing teams will be Germany and Spain with a probability of 13.85%, where Germany will win with a probability of 55%. We can also compute the chance for all teams to reach the quarter-final, the semi-final or the final (Figure 2).

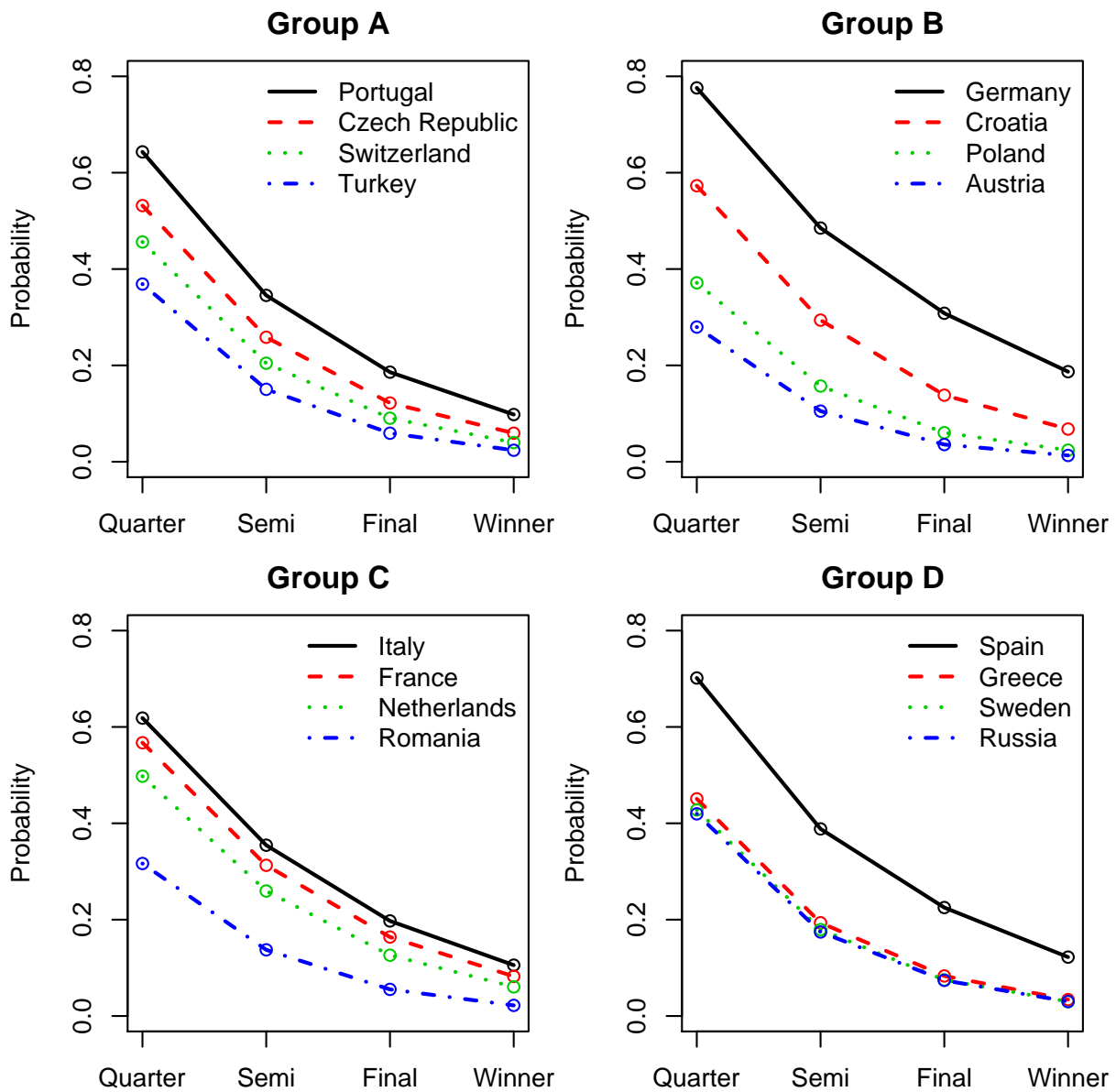


Figure 2: Simulated probabilities (from 100,000 simulated tournaments) to reach the quarter-final, the semi-final, the final and to win the European championship 2008 for each of the 16 teams by simulating the tournament matches based on the pairwise winning probabilities (see Figure 1).

Figure 2 shows that the performance curves of the teams in group B are a long way away from each other. Team Germany has clearly the highest chance to “survive” during the course of the tournament, e.g., with a probability of almost 80% for reaching the quarter-final. The comparison of the curves in group D show that this group has a clear favorite, named team Spain. All other teams of group D have rather similar chances to be promoted to the second round (and also for reaching the semi-final and final) with rather low probabilities to win the tournament for all three teams. The performance curves of the teams in group A and group C are nearly parallel without a clear favorite (as in the other two groups). Furthermore, the expected runner-ups from these groups (in the first round) have relatively low chances

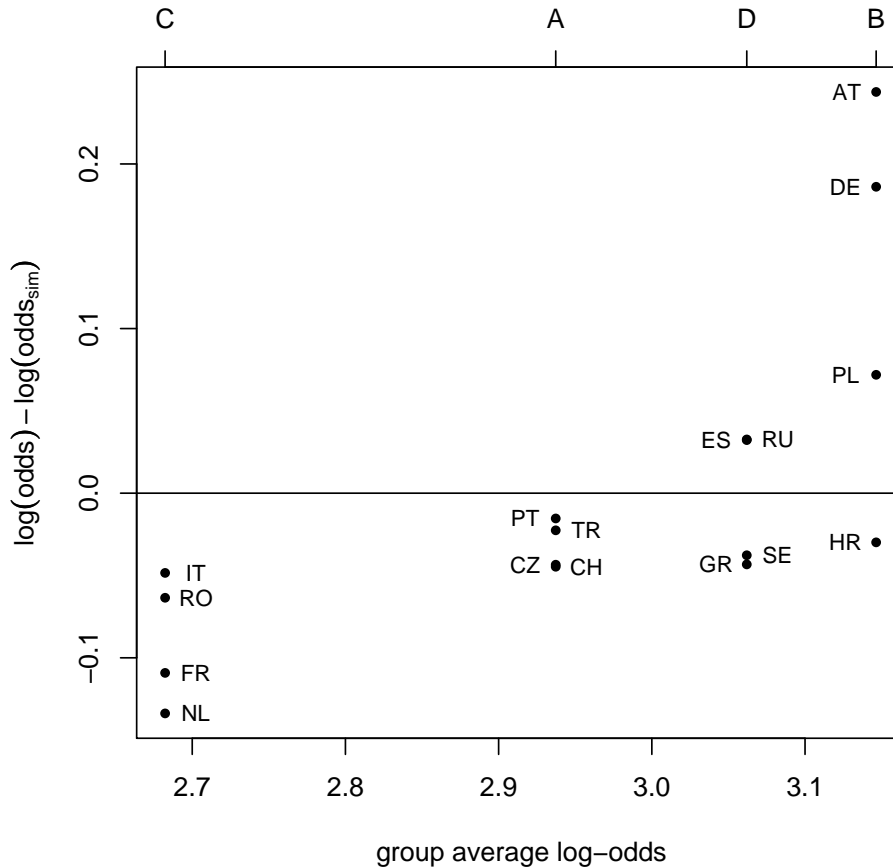


Figure 3: Group effects by comparing the transformed simulated probabilities (in log-odds) of Table 3 and the consensus log-odds of Table 1. For weighting the groups the average log-odds of the group members are shown on the x-axis.

of beating the clear favorites from the other two groups in the resulting quarter-finals.

3.5 Group effects

As shown above the simulated probabilities (Table 3) and the transformed consensus probabilities (Table 1) are not identical. The reason for that is that not every team plays against every other team in this tournament. Of course, this group effect is included in the bookmakers odds (as the group classification is already known) and we attempt to recover it from the consensus rating by using the simulation approach discussed above.

Therefore we use the average log-odds of the group members as group weights (Figure 3). In addition Figure 3 compares the transformed simulated probabilities (in log-odds) of Table 3 and the consensus log-odds of Table 1.

Although our favorite for the title Germany is a group member of group B, group B is according to the log-odds level of the groups (average log-odds of all group members, shown on the x-axis of Figure 3) the

“easiest” group, i.e., the groups with the weakest teams in total (with the highest average log-odds of 3.15). That leads to the conclusion that the other three teams of group B are rather weak. The “group of death”, group C is with an average log-odds of 2.68 clearly the group with the strongest teams. Group A and D are rather balanced groups.

The differences of the the transformed simulated probabilities and the consensus log-odds (shown on the y-axis of Figure 3) show which team profits of the allocated group and which team suffer from the allocated group. Team Austria is therefore the biggest winner of being in the weakest group, i.e., Austria’s probability to win the championship increases most due to this group allocation than playing against each other team. Also team Germany and Poland profit by being member of group B. The only “loser” of this group allocation (group B) is team Croatia. In group C all teams suffer from the fact that they are group members of the “group of death”, whereas the strongest (Italy) and the weakest team (Romania) suffers least. In the balanced groups group A and D the teams profit or suffer not really from their group allocation.

3.6 Rater information

Throughout the paper, we focus on the participating teams of the EURO 2008, but in addition, we can also derive information about the raters, the bookmakers.

By computing the team’s probabilities to win the tournament in the pre-process, the over-rounds for the bookmaker can be estimated. The over-rounds of the 45 bookmakers ($1/\alpha$) are between 0% and 16%, where except one, all other bookmakers have a over-round of at least 1% with a median over-round of 7.4%.

By modeling the relationship between the bookmakers odds and the “true” strength of a team (Section 2.2) we can also model the rating bias $\hat{\mu}_j$ and the variance of the rating error $\hat{\sigma}_j^2$ of bookmaker j for team i . The rating bias $\hat{\mu}_j$ differ between -0.047 and 0.057 on the score scale (in log-odds), where a negative bias means that the bookmaker evaluates the team better than the average rating and a positive bias means that the bookmaker rates the team worse than the average rating. The standard deviation of the rating error $\hat{\sigma}_j$ is between 0.039 and 0.174 on the score scale. In general, the rating bias and the over-round have a negative relationship, i.e., that bookmakers who rate the team rather too good (low odds) have rather higher over-rounds and bookmakers who assign the teams with rather high odds have rather low over-rounds.

4 Discussion

This paper introduces a novel method for rating participants of a sports tournament and predicting the winner based on bookmakers odds for winning the tournament. From theses odds, interpreted as the bookmakers’ rating of the participants’ expected strengths, we derive a consensus rating using a random-effects model with team-specific random effect and bookmaker-specific fixed effect. Furthermore, we suggest to assess such a rating by computing pairwise winning probabilities and simulating the full tournament. Such an approach can capture group and play-off effects and yields more granular information about the probability for each participant to “survive” during the tournament.

In particular, we use this method to predict the winner of the EURO 2008: Team Germany. In addition to the simple prediction, our method quantifies the result by providing winning probabilities (e.g., about 19% for Germany and 12% for Spain, the second-best team), showing that there is a clear tendency for a winner but that the result is by no means certain. Luckily for all soccer fans, this is, like all other sports, a game and cannot be truly predicted using rational strategies and statistical methods.

Computational details

For all calculations we use the R system (version 2.6.1) for statistical computing (R Development Core Team, 2007). To estimate the parameters of the one-factor random-effects model (e.g., Pinheiro and Bates, 2000), modeling the relationship between the bookmakers odds and the “true” strength of a team we use the R package nlme version 3.1-86 (Pinheiro *et al.*, 2007). For a useful output of the pairwise probabilities (Figure 1) we implement a relationship plot function using the strucplot framework according to Meyer *et al.* (2006) and the corresponding strucplot function of the R package vcd version 1.0-6 (Meyer *et al.*, 2007).

References

- Advanced Satellite Consulting Ltd (2008). “The World Football Elo Rating System.” [Online; accessed 2008-04-21], URL <http://www.eloratings.net/>.
- Fédération Internationale de Football Association (2008). “FIFA/Coca-Cola World Ranking.” [Online; accessed 2008-04-21], URL <http://www.fifa.com/worldfootball/ranking/lastranking/gender=m/fullranking.html>.
- Meyer D, Zeileis A, Hornik K (2006). “The Strucplot Framework: Visualizing Multi-Way Contingency Tables with vcd.” *Journal of Statistical Software*, **17**(3), 1–48. URL <http://www.jstatsoft.org/v17/i03/>.
- Meyer D, Zeileis A, Hornik K (2007). *vcd: Visualizing Categorical Data*. R package version 1.0-6, URL <http://cran.r-project.org/package=vcd>.
- Pinheiro J, Bates D (2000). *Mixed-Effects Models in S and S-PLUS*. Statistics and Computing. Springer-Verlag New York, New York, USA. ISBN 0-387-98957-9.
- Pinheiro J, Bates D, DebRoy S, Sarkar D (2007). *nlme: Linear and Nonlinear Mixed Effects Models*. R package version 3.1-86, URL <http://cran.r-project.org/package=nlme>.
- R Development Core Team (2007). *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/eBusiness/rzb_template1/1023296711504-1023296711595_1024688700058-471732017160149076-NA-NA-DE.html.
- UBS Wealth Management Research Switzerland (2008). “European Champions for 2008 Will Be...” [Online; accessed 2008-04-21], URL http://www.ubs.com/1/e/ubs_ch/wealth_management_switzerland/euro.html.
- Wikipedia (2008). “Odds — Wikipedia, The Free Encyclopedia.” [Online; accessed 2008-04-21], URL <http://en.wikipedia.org/w/index.php?title=Odds&oldid=204179204>.

A Bookmakers odds

	CZ	PT	CH	TR	AT	HR	DE	PL
bwin	15.00	8.50	23.00	41.00	81.00	14.00	4.75	41.00
X888	13.00	9.00	15.00	41.00	81.00	13.00	5.00	29.00
bet365	13.00	8.50	21.00	34.00	101.00	12.00	4.50	41.00
betdirect	15.00	8.00	17.00	34.00	101.00	13.00	5.00	51.00
bet1128	17.00	9.00	23.00	37.00	91.00	14.00	5.50	42.00
betChronicle	14.00	8.30	27.00	50.00	104.00	14.00	5.00	43.00
betfred	13.00	8.00	21.00	41.00	101.00	13.00	5.00	34.00
betinternet	15.00	8.50	21.00	41.00	101.00	13.00	5.00	41.00
better	15.00	8.00	17.00	34.00	101.00	13.00	5.00	51.00
bluesq	13.00	9.00	15.00	41.00	81.00	13.00	5.00	29.00
boylesports	15.00	8.00	21.00	34.00	81.00	13.00	5.00	41.00
canbet	15.00	8.00	26.00	41.00	101.00	13.00	4.75	46.00
centrebet	15.00	8.50	23.00	41.00	101.00	11.00	5.00	29.00
coral	15.00	9.00	23.00	34.00	81.00	13.00	4.50	51.00
ladbrokes	15.00	9.00	21.00	34.00	81.00	13.00	5.00	34.00
lasseters	15.00	8.50	21.00	41.00	101.00	13.00	5.00	41.00
paddypower	13.00	8.50	19.00	34.00	67.00	13.00	5.00	34.00
pagebet	13.00	8.50	23.00	41.00	101.00	12.00	5.00	51.00
partybets	11.00	8.50	19.00	29.00	81.00	11.00	5.50	34.00
skybet	17.00	8.00	21.00	34.00	67.00	13.00	5.00	41.00
sportingbet	15.00	8.50	21.00	41.00	101.00	13.00	4.50	51.00
stanjames	15.00	8.00	17.00	34.00	101.00	13.00	5.00	51.00
totesport	15.00	9.00	21.00	34.00	101.00	13.00	4.50	41.00
vcbet	13.00	8.50	17.00	34.00	81.00	11.00	4.50	41.00
hill	12.00	8.00	21.00	34.00	81.00	11.00	5.00	41.00
pinaleports	15.33	8.97	23.46	41.40	83.82	12.24	5.10	41.40
expekt	13.00	9.00	21.00	29.00	67.00	12.00	5.00	29.00
gamebookers	11.00	8.50	19.00	29.00	81.00	11.00	5.00	34.00
bet.at.home	15.00	8.50	22.00	34.00	90.00	12.00	5.00	40.00
gera	13.00	8.50	21.00	34.00	81.00	12.00	4.75	39.00
sunmarker	15.00	8.00	25.00	45.00	100.00	13.00	5.00	45.00
noxwin	15.00	8.00	25.00	45.00	100.00	13.00	5.00	45.00
betway	13.00	8.50	21.00	34.00	81.00	11.00	5.00	34.00
betsafe	13.00	8.50	22.00	40.00	100.00	12.00	5.00	45.00
betboo	15.00	8.50	21.00	34.00	81.00	13.00	4.25	41.00
intertops	15.00	8.50	23.00	51.00	101.00	12.00	4.00	41.00
unibet	15.00	8.50	27.00	45.00	100.00	12.50	5.00	40.00
mybet	13.00	8.00	26.00	41.00	81.00	13.00	5.00	41.00
betsson	15.00	9.00	22.00	35.00	100.00	12.00	5.00	40.00
nordicbet	15.00	9.00	25.00	30.00	75.00	12.00	5.40	40.00
digibet	15.00	8.00	27.00	50.00	100.00	13.00	5.00	50.00
betclick	16.00	8.00	20.00	40.00	80.00	14.00	5.00	40.00
admiralbet	15.00	8.00	20.00	20.00	100.00	12.00	5.00	40.00
interwetten	16.00	10.00	20.00	30.00	100.00	12.00	5.00	50.00
bet24	15.00	9.25	25.00	40.00	100.00	12.50	5.35	50.00

Table 4: Published long term odds of 45 international bookmakers for the teams of group A (Czech Republic, Portugal, Switzerland, and Turkey) and of group B (Austria, Croatia, Germany, and Poland) for winning the EURO 2008 (sources: Webpages of the bookmakers, online: accessed 2008-04-21).

	FR	IT	NL	RO	GR	RU	ES	SE
bwin	10.00	7.00	14.00	34.00	26.00	29.00	7.00	29.00
X888	8.50	8.00	12.00	41.00	26.00	29.00	7.50	26.00
bet365	9.00	8.00	13.00	41.00	23.00	34.00	7.00	29.00
betdirect	9.00	7.00	11.00	41.00	21.00	21.00	7.00	26.00
bet1128	8.50	7.50	12.00	41.00	25.00	35.00	7.50	28.00
betChronicle	9.20	7.90	13.50	39.00	29.00	41.00	7.10	31.00
betfred	9.00	8.00	13.00	51.00	26.00	34.00	7.00	34.00
betinternet	9.50	8.00	13.00	41.00	26.00	34.00	7.00	29.00
better	9.00	7.00	11.00	41.00	21.00	21.00	7.00	26.00
bluesq	8.50	8.00	12.00	41.00	26.00	29.00	7.50	26.00
boylesports	10.00	7.50	15.00	29.00	26.00	26.00	7.00	34.00
canbet	9.00	6.50	12.00	26.00	26.00	31.00	6.00	34.00
centrebet	10.00	8.00	15.00	51.00	26.00	34.00	7.20	34.00
coral	8.00	7.00	13.00	29.00	26.00	34.00	7.00	29.00
ladbrokes	9.00	7.00	13.00	41.00	26.00	29.00	7.00	26.00
lasseters	9.00	7.00	13.00	34.00	26.00	31.00	7.00	31.00
paddypower	9.00	8.00	13.00	41.00	26.00	29.00	7.00	26.00
pagebet	9.00	8.00	13.00	41.00	26.00	29.00	7.00	29.00
partybets	10.00	7.00	10.00	29.00	21.00	26.00	7.00	29.00
skybet	9.00	8.00	11.00	51.00	26.00	26.00	7.00	26.00
sportingbet	10.00	7.00	11.00	34.00	21.00	34.00	6.50	34.00
stanjames	9.00	7.00	11.00	41.00	21.00	21.00	7.00	26.00
totesport	9.00	7.50	13.00	51.00	26.00	34.00	7.00	26.00
vcbet	9.00	8.00	13.00	34.00	26.00	26.00	6.50	26.00
hill	10.00	8.00	13.00	41.00	26.00	26.00	7.00	26.00
pinalesports	10.13	7.37	12.35	41.40	29.88	36.11	7.17	35.57
expekt	9.00	7.50	10.00	34.00	26.00	29.00	7.00	26.00
gamebookers	10.00	7.00	10.00	29.00	21.00	26.00	7.00	29.00
bet.at.home	10.00	8.00	14.00	40.00	25.00	30.00	7.00	30.00
gera	10.00	7.00	13.00	34.00	23.00	29.00	7.00	29.00
sunmarker	9.00	7.50	13.00	40.00	25.00	45.00	6.00	30.00
noxwin	9.00	7.50	13.00	40.00	25.00	45.00	6.00	30.00
betway	10.00	8.00	13.00	34.00	23.00	26.00	7.00	26.00
betsafe	10.00	8.00	13.00	40.00	26.00	32.00	7.00	32.00
betboo	8.00	6.50	11.00	34.00	21.00	31.00	6.50	26.00
intertops	9.00	8.00	14.00	41.00	26.00	34.00	6.50	29.00
unibet	10.00	8.00	12.50	45.00	33.00	35.00	7.00	33.00
mybet	9.00	7.50	13.00	34.00	26.00	29.00	6.80	29.00
betsson	8.00	8.00	10.00	50.00	30.00	30.00	8.00	25.00
nordicbet	10.00	5.50	12.00	40.00	25.00	30.00	7.00	25.00
digibet	9.00	7.50	13.50	50.00	27.00	40.00	6.50	33.00
betclick	10.00	8.00	14.00	40.00	25.00	35.00	7.00	30.00
admiralbet	10.00	8.00	12.00	40.00	20.00	30.00	8.00	30.00
interwetten	7.50	7.50	12.00	50.00	30.00	30.00	7.50	30.00
bet24	10.00	7.50	12.50	50.00	30.00	35.00	7.00	35.00

Table 5: Published long term odds of 45 international bookmakers for the teams of group C (France, Italy, Netherlands, and Romania) and of group D (Greece, Russia, Spain, Sweden) for winning the EURO 2008 (sources: Webpages of the bookmakers, online: accessed 2008-04-21).