Forecasting sports tournaments by ratings of (prob)abilities

Achim Zeileis, Christoph Leitner, Kurt Hornik

http://eeecon.uibk.ac.at/~zeileis/
UEFA Euro 2016 prediction

Round of 16

25 June - Saint-Étienne
- Switzerland 1 (4)
- Poland (pen.) 1 (5)
- Portugal (a.e.t.) 1

25 June - Lens
- Croatia 0
- Portugal (pen.) 1 (5)

25 June - Paris
- Wales 1
- Northern Ireland 0
- Belgium 4

30 June - Marseille
- Poland 1 (3)
- Portugal (pen.) 1 (5)

Quarter-finals

1 July - Villeneuve-d'Ascq
- Wales 3
- Belgium 1

2 July - Bordeaux
- Germany (pen.) 1 (6)
- Italy 1 (5)

3 July - Saint-Denis
- France 5
- Iceland 2

Semi-finals

6 July - Lyon
- Portugal 2
- Wales 0

7 July - Marseille
- Germany 0
- France 2

Final

10 July - Saint-Denis
- Portugal (a.e.t.) 1
- France 0

Source: UEFA, Wikipedia
Tournament forecast based on bookmakers odds.

Main results: France and Germany are the top favorites with winning probabilities of 21.5% and 20.1%, respectively.

Top favorites are most likely to meet in the semifinal with odds very slightly in favor of France (50.5% winning probability).
• All favorites “survive” the group stage.
• But: Spain and England blow the chance of winning their respective groups.
• Austria is eliminated after disappointing performances.
England surprisingly loses to Iceland.

Spain loses the “replay” of the Euro 2012 final against Italy.
- Wales surprisingly beats Belgium.
- After a strong tournament Iceland clearly loses to France.
For the first and only time Portugal wins a match after 90 minutes.

In the match of the top favorites France beats Germany despite a strong performance of the world champion.
Host France fails to seal the victory in normal time and loses to Portugal after extra time.
Bookmakers odds

Source: williamhill.com, bwin.com
Bookmakers odds: Motivation

Forecasts of sports events:

- Increasing interest in forecasting of competitive sports events due to growing popularity of online sports betting.
- Forecasts often based on ratings or rankings of competitors’ ability/strength.

In football:

- Elo rating.
  - Aims to capture relative strength of competitors yielding probabilities for pairwise comparisons.
  - Originally developed for chess.
- FIFA rating.
  - Official ranking, used for seeding tournaments.
  - Often criticized for not capturing current strengths well.
Alternately: Employ bookmakers odds for winning a competition.

- Bookmakers are “experts” with monetary incentives to rate competitors correctly. Setting odds too high/low yields less profits.
- Prospective in nature: Bookmakers factor not only the competitors abilities into their odds but also tournament draws/seedings, home advantages, recent events such as injuries, etc.
- Statistical “post-processing” needed to derive winning probabilities and underlying abilities.
Bookmakers odds: Overround adjustment

**Odds:** In statistics, the ratio of the probabilities for winning/losing, e.g.
- Even odds are “50:50” (= 1).
- Odds of 4 correspond to probabilities $4/5 = 80\%$ vs. $1/5 = 20\%$.

**Quoted odds:** In sports betting, the payout for a stake of 1.

This is not an honest judgment of winning chances due to inclusion of a profit margin known as “overround”.

$$quoted\ odds_i = odds_i \cdot \delta + 1,$$

- where $odds_i$ is the bookmaker’s “true” judgment of the odds for competitor $i$,
- $\delta$ is the bookmaker’s payout proportion (overround: $1 - \delta$),
- and $+1$ is the stake.
**Bookmakers odds: Overround adjustment**

**Winning probabilities:** The adjusted odds $odds_i$ then corresponding to the odds of competitor $i$ for losing the tournament. They can be easily transformed to the corresponding winning probability

$$p_i = 1 - \frac{odds_i}{1 + odds_i}.$$

**Determining the overround:** Assuming that a bookmaker’s overround is constant across competitors, it can be determined by requiring that the winning probabilities of all competitors (here: all 24 teams) sum to 1: $\sum_i p_i = 1$. 

Illustration: UEFA Euro 2016 rating for France by bookmaker bwin.

- Bookmaker bwin pays 4.33 for a stake of 1 set on a victory of France, i.e., a profit of 3.33.
- The overround implied by bwin’s quoted odds for all 24 teams in the tournament is 14.4%.
- Thus, bwin’s implied odds for France are:
  \[ 3.89 = \frac{4.33 - 1}{1 - 0.144} \]
  i.e., it is about four times more likely that France loses vs. wins.
- The corresponding winning probability for France is 20.4%.
Data processing:

- Quoted odds from 19 online bookmakers.
- Computed overrounds $1 - \delta_b$ individually for each bookmaker $b = 1, \ldots, 19$ by unity sum restriction across teams $i = 1, \ldots, 24$.
- Median overround is $15.1\%$.
- Yields overround-adjusted and transformed winning probabilities $\rho_{i,b}$ for each team $i$ and bookmaker $b$. 
Modeling consensus and agreement
Modeling consensus and agreement

**Goal:** Get consensus probabilities by aggregation across bookmakers.

**Strategy:**
- Employ statistical model assuming some latent consensus probability \( p_i \) for team \( i \) along deviations \( \varepsilon_{i,b} \).
- Additive model is plausible on suitable scale, e.g., logit or probit.
- Logit is more natural here, as it corresponds to log-odds.
- Methodology can also be used for consensus ratings of default probability in credit risk rating of bank \( b \) for firm \( i \).

**Model:** Bookmaker consensus model

\[
\text{logit}(p_{i,b}) = \text{logit}(p_i) + \varepsilon_{i,b},
\]

where further effects could be included, e.g., group effects in consensus logits or bookmaker-specific bias and variance in \( \varepsilon_{i,b} \).
Modeling consensus and agreement

Here:

- Simple fixed-effects model with zero-mean deviations.
- Consensus logits are simply team-specific means across bookmakers:
  \[ \hat{\logit}(p_i) = \frac{1}{19} \sum_{b=1}^{19} \logit(p_{i,b}). \]
- Consensus winning probabilities are obtained by transforming back to the probability scale:
  \[ \hat{p}_i = \logit^{-1}\left(\hat{\logit}(p_i)\right). \]
- Model captures 97.9% of the variance in \( \logit(p_{i,b}) \) and the associated estimated standard error is 0.204.
# Modeling consensus and agreement

<table>
<thead>
<tr>
<th>Team</th>
<th>FIFA code</th>
<th>Probability</th>
<th>Log-odds</th>
<th>Log-ability</th>
<th>Group</th>
</tr>
</thead>
<tbody>
<tr>
<td>France</td>
<td>FRA</td>
<td>21.5</td>
<td>-1.298</td>
<td>-1.748</td>
<td>A</td>
</tr>
<tr>
<td>Germany</td>
<td>GER</td>
<td>20.1</td>
<td>-1.379</td>
<td>-1.766</td>
<td>C</td>
</tr>
<tr>
<td>Spain</td>
<td>ESP</td>
<td>13.7</td>
<td>-1.840</td>
<td>-2.001</td>
<td>D</td>
</tr>
<tr>
<td>England</td>
<td>ENG</td>
<td>9.2</td>
<td>-2.290</td>
<td>-2.209</td>
<td>B</td>
</tr>
<tr>
<td>Belgium</td>
<td>BEL</td>
<td>7.7</td>
<td>-2.489</td>
<td>-2.261</td>
<td>E</td>
</tr>
<tr>
<td>Italy</td>
<td>ITA</td>
<td>5.1</td>
<td>-2.932</td>
<td>-2.393</td>
<td>E</td>
</tr>
<tr>
<td>Portugal</td>
<td>POR</td>
<td>4.1</td>
<td>-3.146</td>
<td>-2.538</td>
<td>F</td>
</tr>
<tr>
<td>Croatia</td>
<td>CRO</td>
<td>2.9</td>
<td>-3.508</td>
<td>-2.633</td>
<td>D</td>
</tr>
<tr>
<td>Austria</td>
<td>AUT</td>
<td>2.3</td>
<td>-3.751</td>
<td>-2.771</td>
<td>F</td>
</tr>
<tr>
<td>Poland</td>
<td>POL</td>
<td>1.7</td>
<td>-4.038</td>
<td>-2.892</td>
<td>C</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Abilities and tournament simulations

\[ \Pr(i \text{ beats } j) = \pi_{i,j} \]

\[ \pi_{i,j} = \frac{\text{ability}_i}{\text{ability}_i + \text{ability}_j} \]

Abilities and tournament simulations

Further questions:

- What are the likely courses of the tournament that lead to these bookmaker consensus winning probabilities?
- Is the team with the highest probability also the strongest team?
- What are the winning probabilities for all possible matches?

Motivation:

- Tournament draw might favor some teams, e.g., France was drawn in a group with two weak teams (Romania and Albania).
- Tournament schedule was known to bookmakers and hence factored into their quoted odds.
- Can abilities (or strengths) of the teams be obtained, adjusting for such tournament effects?
Abilities and tournament simulations

**Answer:** Yes, an approximate solution can be found by simulation when

- adopting a standard model for paired comparisons (i.e., matches),
- assuming that the abilities do not change over the tournament.

**Model:** Bradley-Terry model for winning/losing in a paired comparison of team $i$ and team $j$.

$$\Pr(i \text{ beats } j) = \pi_{i,j} = \frac{ability_i}{ability_i + ability_j}.$$
Abilities and tournament simulations

“Reverse” simulation:

- If the team-specific ability $i$ were known, pairwise probabilities $\pi_{i,j}$ could be computed.
- Given $\pi_{i,j}$ the whole tournament can be simulated (assuming abilities do not change and ignoring possible draws during the group stage).
- Using “many” simulations (here: 100,000) of the tournament, the empirical relative frequencies $\tilde{p}_i$ of each team $i$ winning the tournament can be determined.
- Choose ability $i$ for $i = 1, \ldots, 24$ such that the simulated winning probabilities $\tilde{p}_i$ approximately match the consensus winning probabilities $\hat{p}_i$.
- Found by simple iterative local search starting from log-odds.
## Abilities and paired comparisons

A heatmap showing the paired comparisons between teams across various abilities. Each cell represents the comparison score between two teams, with darker shades indicating higher scores. The horizontal axis represents Team j, while the vertical axis represents Team i.
Tournament simulations: Survival curves

Group A

- FRA
- SUI
- ROU
- ALB

Group B

- ENG
- RUS
- WAL
- SVK
Tournament simulations: Survival curves

Group C

- GER
- POL
- UKR
- NIR

Group D

- ESP
- CRO
- TUR
- CZE
Tournament simulations: Survival curves

**Group E**

- Probability (%)
- Round of 16
- Quarter
- Semi
- Final
- Winner

- BEL
- ITA
- SWE
- IRL

**Group F**

- Probability (%)
- Round of 16
- Quarter
- Semi
- Final
- Winner

- POR
- AUT
- ISL
- HUN
Outcome verification
Outcome verification

**Question:** Was the forecast any good?

- Ex post the low predicted winning probability for Portugal (4.1%) seems wrong.
- However, consider that they indirectly profited from Spain’s and England’s poor performances in the last group stage games.
- And they only won 1 out of 7 games in normal time.
- Even in the final Gignac might as well have scored a goal instead of hitting the post in minute 92... 

**Problems:**

- Just a single observation of the tournament and at most one observation of each paired comparison.
- Hard to distinguish between occurrence of an un- (or less) likely outcome and systematic errors in the predicted (prob)abilities.
Outcome verification

Possible approaches:

- Compare forecasts with the observed tournament ranking (1 POR, 2 FRA, 3.5 WAL, 3.5 GER, ...).
- Benchmark against Elo and FIFA ratings.
- Note that the Elo rating also implies ability scores based on which pairwise probabilities and “forward” simulation of tournament can be computed:
  
  $ability_{Elo,i} = 10^{Elo_i/400}$.

- Check whether pairwise probabilities roughly match empirical proportions from clusters of matches.
Outcome verification: Ranking

Spearman rank correlation of observed tournament ranking with bookmaker consensus model (BCM) as well as FIFA and Elo ranking:

<table>
<thead>
<tr>
<th>Model Type</th>
<th>Correlation</th>
</tr>
</thead>
<tbody>
<tr>
<td>BCM (Probabilities)</td>
<td>0.523</td>
</tr>
<tr>
<td>BCM (Abilities)</td>
<td>0.436</td>
</tr>
<tr>
<td>Elo (Probabilities)</td>
<td>0.344</td>
</tr>
<tr>
<td>Elo</td>
<td>0.339</td>
</tr>
<tr>
<td>FIFA</td>
<td>0.310</td>
</tr>
</tbody>
</table>
# Outcome verification: BCM pairwise probabilities

<table>
<thead>
<tr>
<th>Winning probability of stronger team (in %)</th>
<th>Win</th>
<th>Draw</th>
<th>Lose</th>
</tr>
</thead>
<tbody>
<tr>
<td>[50, 60]</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(60, 70)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(70, 85]</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The diagram illustrates the probability distribution of outcomes for pairwise comparisons, categorized into 'Win', 'Draw', and 'Lose' for three different ranges of winning probabilities. The x-axis represents the winning probability of the stronger team, ranging from 0.0 to 1.0.
Outcome verification: BCM pairwise probabilities

Winning probability of stronger team (in %)

Win | Draw | Lose
---|------|-----
0.0 | 0.2  | 0.4
0.6 | 0.8  | 1.0

Winning probability of stronger team (in %)

[50,60] | (60,70) | (70,85)
Outcome verification: Elo pairwise probabilities

Winning probability of stronger team (in %)

- [50, 65]
- (65, 80]
- (80, 95]

Win | Draw | Lose
---|------|------
0.0 | 0.2  | 0.4  | 0.6  | 0.8  | 1.0  

Winning probability of stronger team (in %)
Outcome verification: BCM abilities

Relative ability (BCM)

FRA GER ESP ENG BEL ITA POR CRO AUT POL SUI RUS WAL TUR UKR CZE ISL SWE IRL SVK ROU HUN NIR ALB
Outcome verification: Elo abilities

Relative ability (Elo)

<table>
<thead>
<tr>
<th>FRA</th>
<th>GER</th>
<th>ESP</th>
<th>ENG</th>
<th>BEL</th>
<th>ITA</th>
<th>POR</th>
<th>CRO</th>
<th>AUT</th>
<th>POL</th>
<th>SUI</th>
<th>RUS</th>
<th>WAL</th>
<th>TUR</th>
<th>UKR</th>
<th>CZE</th>
<th>ISL</th>
<th>SWE</th>
<th>IRL</th>
<th>SVK</th>
<th>ROU</th>
<th>HUN</th>
<th>NIR</th>
<th>ALB</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Median
Discussion

Summary:

- Expert judgments of bookmakers are a useful information source for probabilistic forecasts of sports tournaments.
- Winning probabilities are obtained by adjustment for overround and averaging on log-odds scale.
- Competitor abilities can be inferred by post-processing based on pairwise-comparison model with “reverse” tournament simulations.
- Approach outperformed Elo and FIFA ratings for the last UEFA Euros and correctly predicted the final 2008 and winner 2012.

Limitations:

- Matches are only assessed in terms of winning/losing, i.e., no goals, draws, or even more details.
- Inherent chance component is substantial and hard to verify.


## Groups A and B

<table>
<thead>
<tr>
<th>Rank</th>
<th>Team</th>
<th>Probability (in %)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>FRA</td>
<td>97.8</td>
</tr>
<tr>
<td>2</td>
<td>SUI</td>
<td>66.9</td>
</tr>
<tr>
<td>3</td>
<td>ALB</td>
<td>39.4</td>
</tr>
<tr>
<td>4</td>
<td>ROU</td>
<td>52.4</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Rank</th>
<th>Team</th>
<th>Probability (in %)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>WAL</td>
<td>61.2</td>
</tr>
<tr>
<td>2</td>
<td>ENG</td>
<td>91.2</td>
</tr>
<tr>
<td>3</td>
<td>SVK</td>
<td>51.7</td>
</tr>
<tr>
<td>4</td>
<td>RUS</td>
<td>64.8</td>
</tr>
</tbody>
</table>
Groups C and D

<table>
<thead>
<tr>
<th>Rank</th>
<th>Team</th>
<th>Probability (in %)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>GER</td>
<td>96.8</td>
</tr>
<tr>
<td>2</td>
<td>POL</td>
<td>66.8</td>
</tr>
<tr>
<td>3</td>
<td>NIR</td>
<td>37.6</td>
</tr>
<tr>
<td>4</td>
<td>UKR</td>
<td>59.9</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Rank</th>
<th>Team</th>
<th>Probability (in %)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>CRO</td>
<td>71.1</td>
</tr>
<tr>
<td>2</td>
<td>ESP</td>
<td>91.7</td>
</tr>
<tr>
<td>3</td>
<td>TUR</td>
<td>55.6</td>
</tr>
<tr>
<td>4</td>
<td>CZE</td>
<td>53.5</td>
</tr>
</tbody>
</table>
Groups E and F

<table>
<thead>
<tr>
<th>Rank</th>
<th>Team</th>
<th>Probability (in %)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>ITA</td>
<td>83.0</td>
</tr>
<tr>
<td>2</td>
<td>BEL</td>
<td>86.9</td>
</tr>
<tr>
<td>3</td>
<td>IRL</td>
<td>47.2</td>
</tr>
<tr>
<td>4</td>
<td>SWE</td>
<td>54.4</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Rank</th>
<th>Team</th>
<th>Probability (in %)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>HUN</td>
<td>47.0</td>
</tr>
<tr>
<td>2</td>
<td>ISL</td>
<td>62.7</td>
</tr>
<tr>
<td>3</td>
<td>POR</td>
<td>84.5</td>
</tr>
<tr>
<td>4</td>
<td>AUT</td>
<td>75.7</td>
</tr>
</tbody>
</table>
## Round of 16

<table>
<thead>
<tr>
<th>Teams</th>
<th>Probability (in %)</th>
<th>Result</th>
</tr>
</thead>
<tbody>
<tr>
<td>POL</td>
<td>50.6</td>
<td>6:5 (pen.)</td>
</tr>
<tr>
<td>SUI</td>
<td></td>
<td></td>
</tr>
<tr>
<td>WAL</td>
<td>61.1</td>
<td>1:0</td>
</tr>
<tr>
<td>NIR</td>
<td></td>
<td></td>
</tr>
<tr>
<td>POR</td>
<td>52.4</td>
<td>1:0 (a.e.t.)</td>
</tr>
<tr>
<td>CRO</td>
<td></td>
<td></td>
</tr>
<tr>
<td>FRA</td>
<td>79.6</td>
<td>2:1</td>
</tr>
<tr>
<td>IRL</td>
<td></td>
<td></td>
</tr>
<tr>
<td>GER</td>
<td>80.2</td>
<td>3:0</td>
</tr>
<tr>
<td>SVK</td>
<td></td>
<td></td>
</tr>
<tr>
<td>BEL</td>
<td>73.9</td>
<td>4:0</td>
</tr>
<tr>
<td>HUN</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ESP</td>
<td>59.7</td>
<td>0:2</td>
</tr>
<tr>
<td>ITA</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ENG</td>
<td>69.1</td>
<td>1:2</td>
</tr>
<tr>
<td>ISL</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
**Quarterfinal, semifinal, final**

<table>
<thead>
<tr>
<th>Teams</th>
<th>Probability (in %)</th>
<th>Result</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Quarterfinal</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>POL</td>
<td>41.2</td>
<td>4:6 (pen.)</td>
</tr>
<tr>
<td>POR</td>
<td></td>
<td></td>
</tr>
<tr>
<td>WAL</td>
<td>33.4</td>
<td>3:1</td>
</tr>
<tr>
<td>BEL</td>
<td></td>
<td></td>
</tr>
<tr>
<td>GER</td>
<td>65.2</td>
<td>7:6 (pen.)</td>
</tr>
<tr>
<td>ITA</td>
<td></td>
<td></td>
</tr>
<tr>
<td>FRA</td>
<td>78.0</td>
<td>5:2</td>
</tr>
<tr>
<td>ISL</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Semifinal</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>POR</td>
<td>60.2</td>
<td>2:0</td>
</tr>
<tr>
<td>WAL</td>
<td></td>
<td></td>
</tr>
<tr>
<td>GER</td>
<td>49.5</td>
<td>0:2</td>
</tr>
<tr>
<td>FRA</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Final</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>POR</td>
<td>31.2</td>
<td>1:0 (a.e.t.)</td>
</tr>
<tr>
<td>FRA</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>