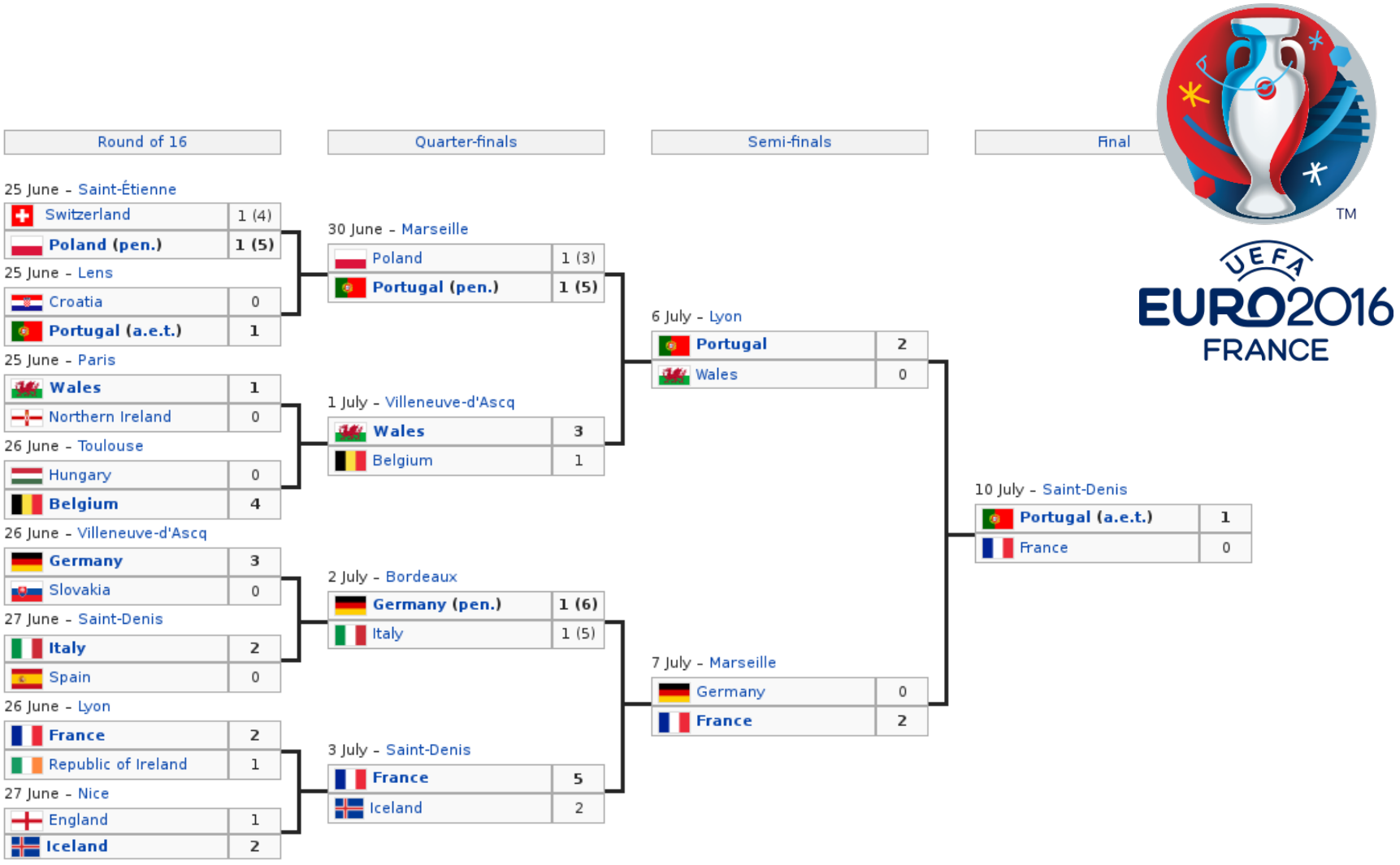
Forecasting sports tournaments by ratings of (prob)abilities

Achim Zeileis, Christoph Leitner, Kurt Hornik

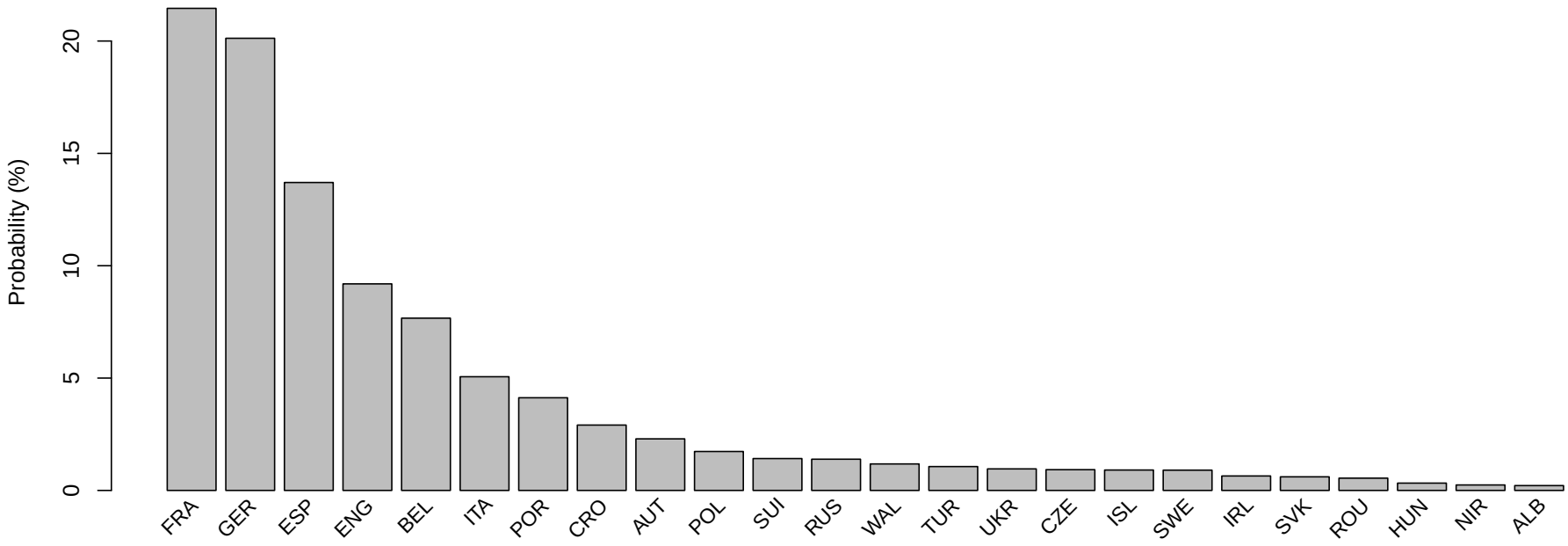
http://eeecon.uibk.ac.at/~zeileis/

UEFA Euro 2016 prediction



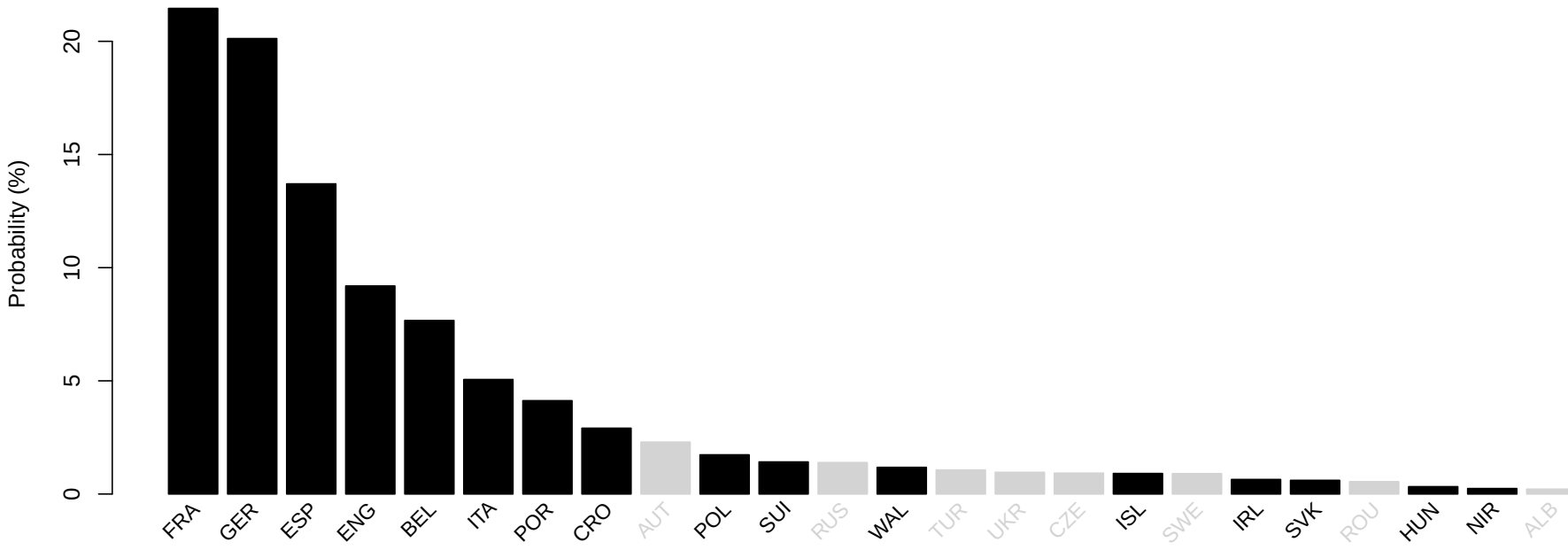
Source: UEFA, Wikipedia

UEFA Euro 2016 prediction



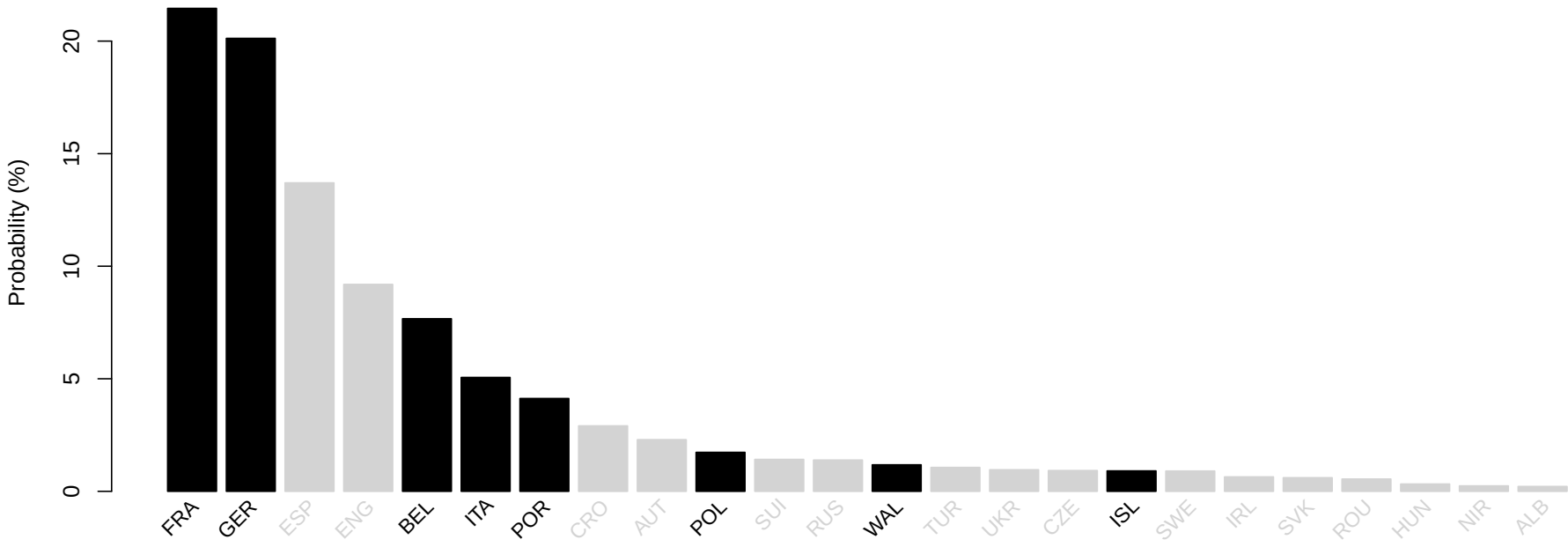
- Tournament forecast based on bookmakers odds.
- Main results: France and Germany are the top favorites with winning probabilities of 21.5% and 20.1%, respectively.
- Top favorites are most likely to meet in the semifinal with odds very slightly in favor of France (50.5% winning probability).

UEFA Euro 2016 tournament



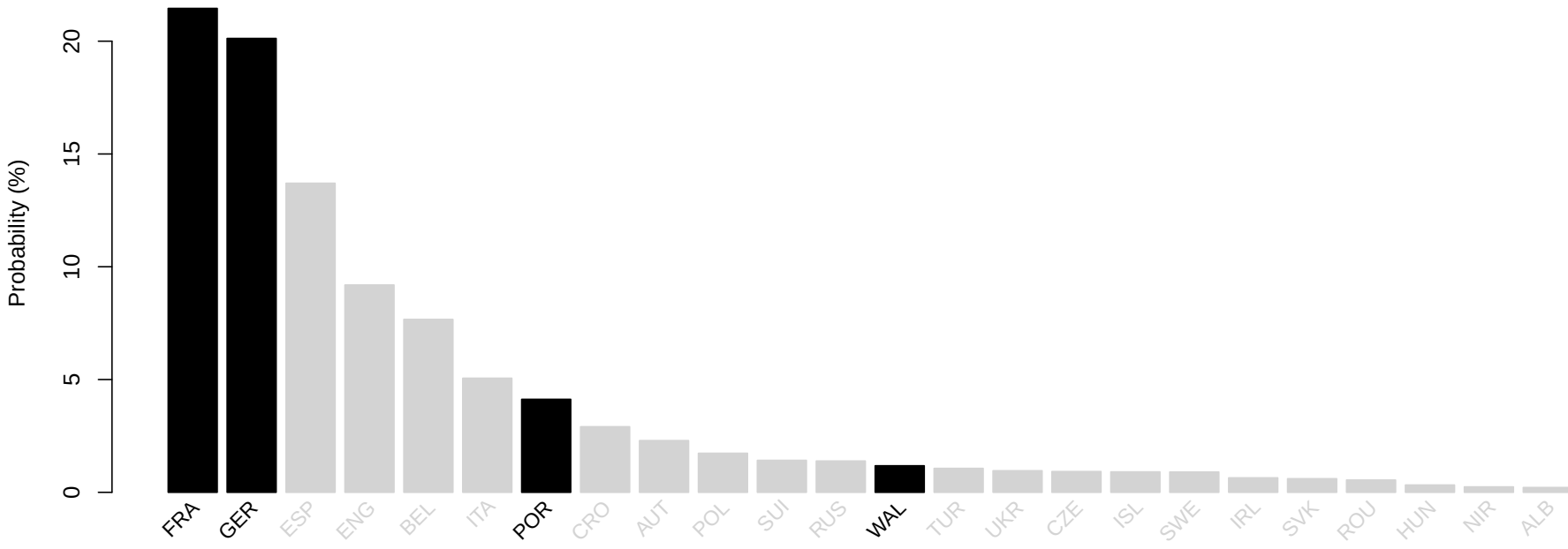
- All favorites “survive” the group stage.
- But: Spain and England blow the chance of winning their respective groups.
- Austria is eliminated after disappointing performances.

UEFA Euro 2016 tournament



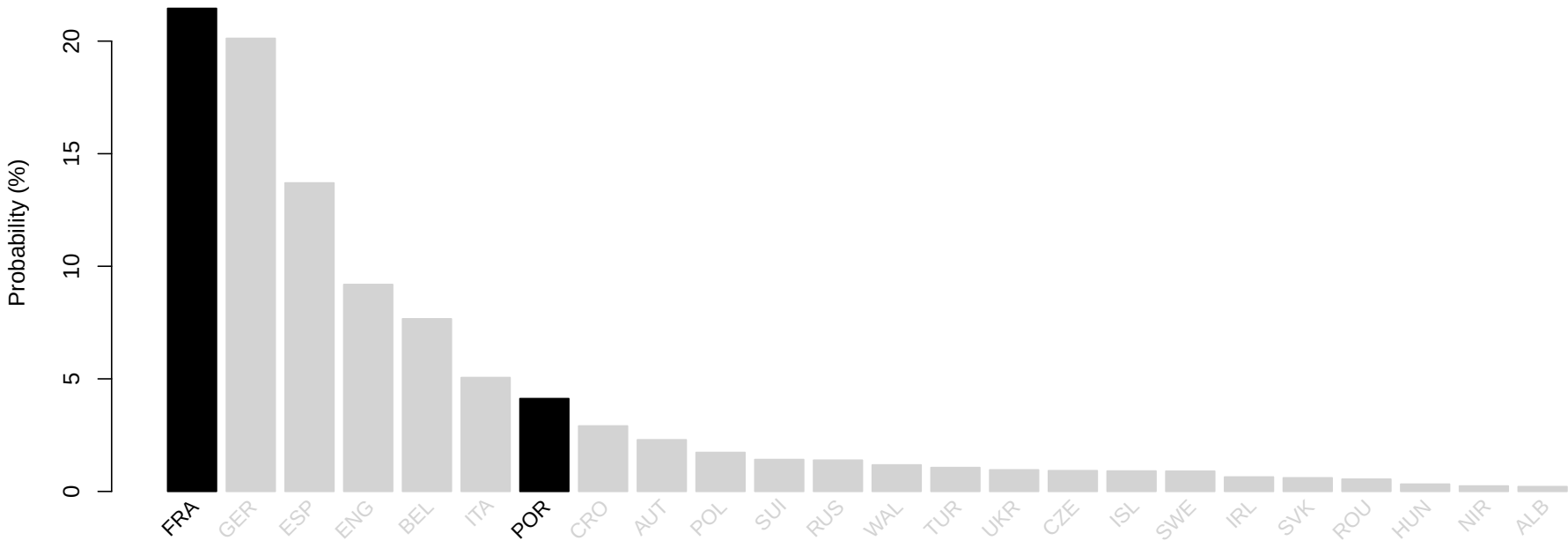
- England surprisingly loses to Iceland.
- Spain loses the “replay” of the Euro 2012 final against Italy.

UEFA Euro 2016 tournament



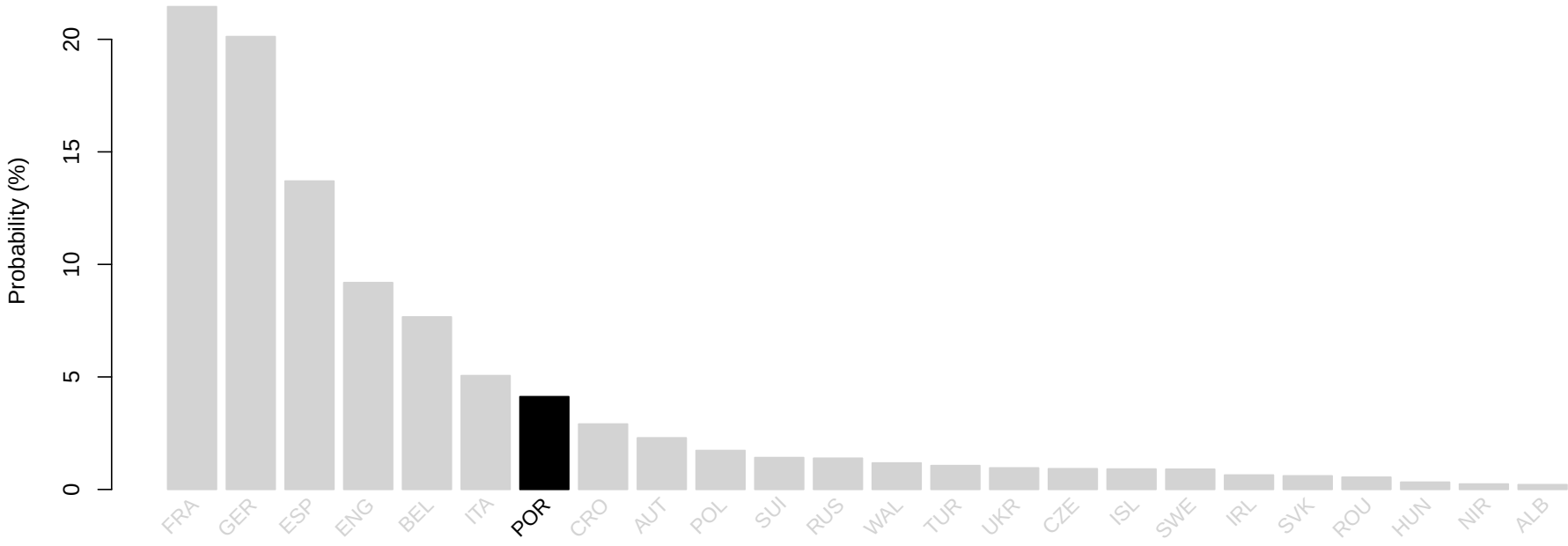
- Wales surprisingly beats Belgium.
- After a strong tournament Iceland clearly loses to France.

UEFA Euro 2016 tournament



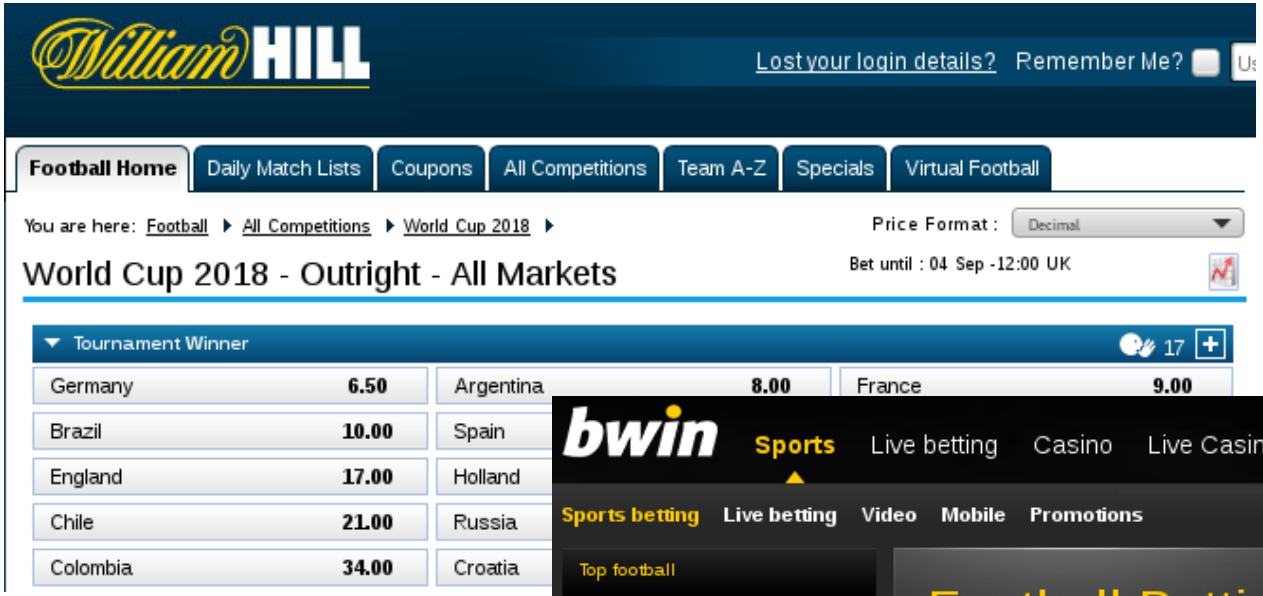
- For the first and only time Portugal wins a match after 90 minutes.
- In the match of the top favorites France beats Germany despite a strong performance of the world champion.

UEFA Euro 2016 tournament



- 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.

Bookmakers odds: Motivation

Alternatively: 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”.

$$\textit{quoted odds}_i \;=\; \textit{odds}_i \cdot \delta + 1,$$

- where $\textit{odds}_i$ is the bookmaker’s “true” judgment of the odds for competitor i,
- δ is the bookmaker’s payout proportion (overround: $1 - \delta$),
- and $+1$ is the stake.

Bookmakers odds: Overround adjustment

Winning probabilities: The adjusted $odds_i$ 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: $\sum_i p_i = 1$.

Bookmakers odds: Overround adjustment

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

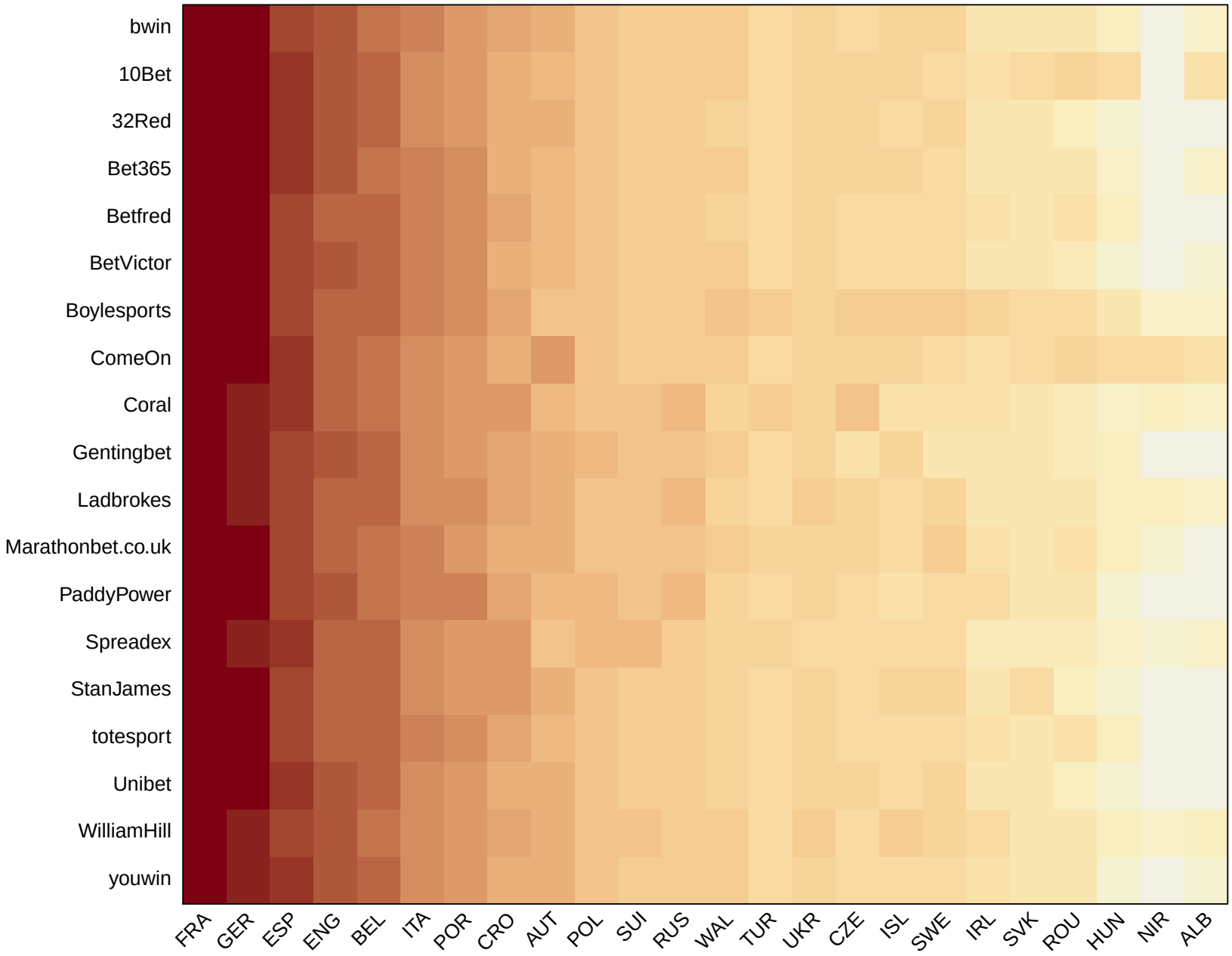
- Bookmaker bwin pays 4.33 for a stake of 1 set on a victory of France, i.e., a profit of 3.33.
- The overround implied by bwin's quoted odds for all 24 teams in the tournament is 14.4%.
- Thus, bwin's implied odds for France are:
  $3.89 = (4.33 - 1)/(1 - 0.144)$, i.e., it is about four times more likely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

Bookmakers odds: UEFA Euro 2016

Data processing:

- Quoted odds from 19 online bookmakers.
- Obtained on 2016-05-22 from `http://www.bwin.com/` and `http://www.oddscomparisons.com/`.
- 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 $p_{i,b}$ 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:

$$\widehat{\text{logit}(p_i)} \;=\; \frac{1}{19}\sum_{b=1}^{19}\text{logit}(p_{i,b}).$$

- Consensus winning probabilities are obtained by transforming back to the probability scale:

$$\hat{p}_i \;=\; \text{logit}^{-1}\left(\widehat{\text{logit}(p_i)}\right).$$

- Model captures 97.9% of the variance in $\text{logit}(p_{i,b})$ and the associated estimated standard error is 0.204.

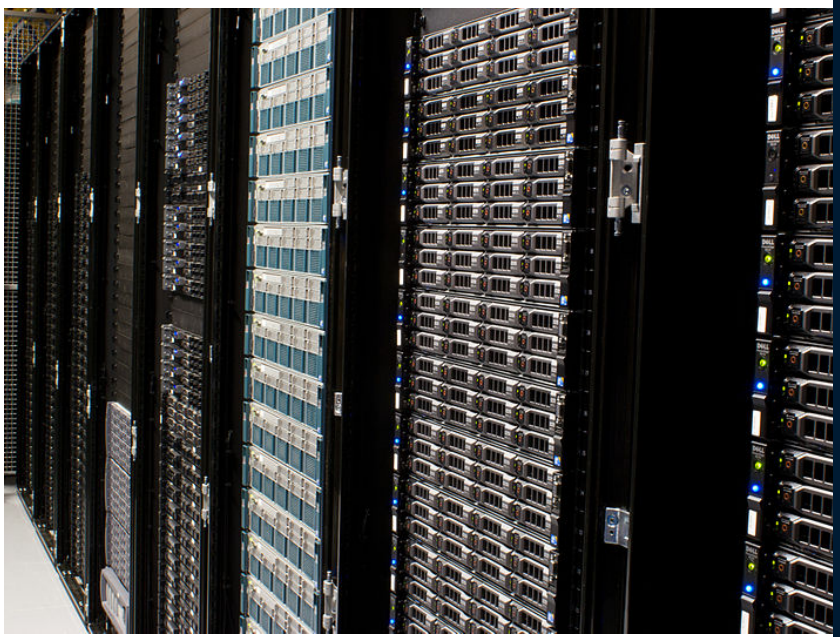
Modeling consensus and agreement

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

Abilities and tournament simulations

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

$$= \frac{ability_i}{ability_i + ability_j}$$



Source: Wikipedia

tournament.R
~/svn/projects/baR/Projects/Euro-2016

```r
sim_log_abilities <- function(logodds, groups,
         start = NULL, n = 100000, rounds = 5,
         loss = function(x, y) mean(abs(x - y), na.rm = TRUE),
         tol = 0.1, maxiter = 100, eps = 1, rate = 0.8,
         cores = NULL, trace = TRUE)
{
  ## main input: winning log-odds
  stopifnot(!is.null(names(logodds)))
  nam <- names(logodds)
  target <- logodds
  if(is.null(start)) start <- logodds
  if(is.null(names(start))) names(start) <- nam

  ## group list
  if(is.null(names(groups))) {
    names(groups) <- nam
  } else {
    groups <- groups[nam]
  }
  groups <- tapply(groups, groups, names)

  ## simulate a full tournament run
  sim1 <- function(log_abilities) {
    simulate_tournament(n = n, probs = get_probs_abilities(exp(log_abilities)),
      groups = groups, cores = cores, rounds = rounds)
  }

  iter <- 1
  if(trace) cat("Start:", start, "\n")
  x <- list()
  y <- list()
  loss_value <- list()
  x[[1]] <- start[names(target)]
  repeat {
    result <- sim1(x[[iter]])
    winner_i <- factor(sapply(result, "[[", "winner"), levels = nam)
    prob_i <- pmax(prop.table(table(winner_i)), 1/n)
    y[[iter]] <- qlogis(prob_i)[names(target)]
    if(trace) {
      cat("* Iteration:", iter, "\n")
      cat("* Log_abilties:", x[[iter]], "\n")
    }
    loss_value[[iter]] <- loss(y[[iter]], target)
    if(trace) cat("Value of the loss function:", round(loss_value[[iter]], 4), "\n")
    if((loss_value[[iter]] < tol) || (iter >= maxiter))
    break
    iter <- iter + 1
    x[[iter]] <- x[[iter-1]] - (y[[iter-1]] - target) / abs(y[[iter-1]] - target) * eps / (
  }
  list(log_abilities = x, result = result, loss_value = loss_value)
}
```

R Tab Width: 8 Ln 208, Col 20 INS

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*.

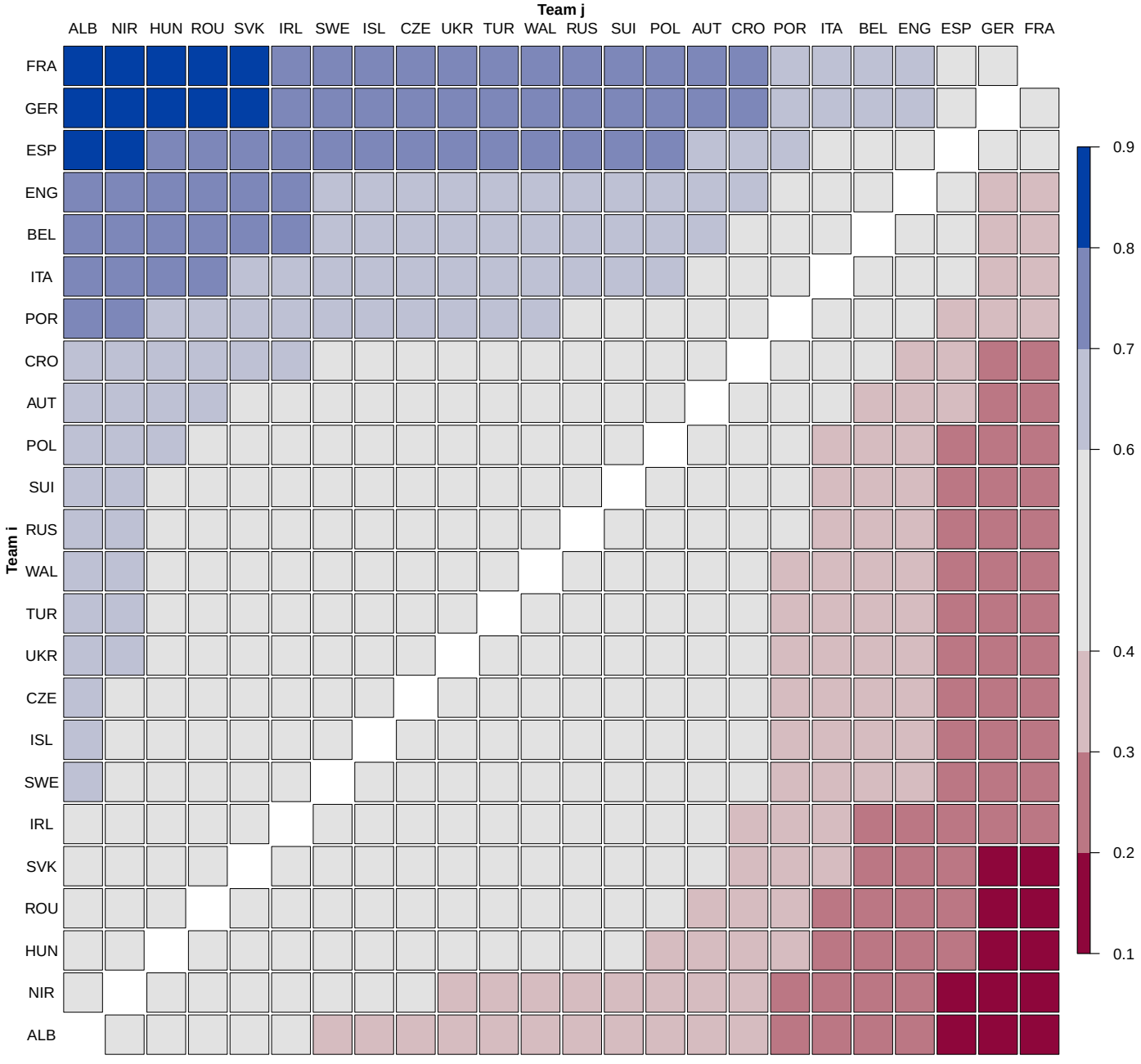
$$\Pr(i \text{ beats } j) \;=\; \pi_{i,j} \;=\; \frac{\mathit{ability}_i}{\mathit{ability}_i + \mathit{ability}_j}.$$

Abilities and tournament simulations

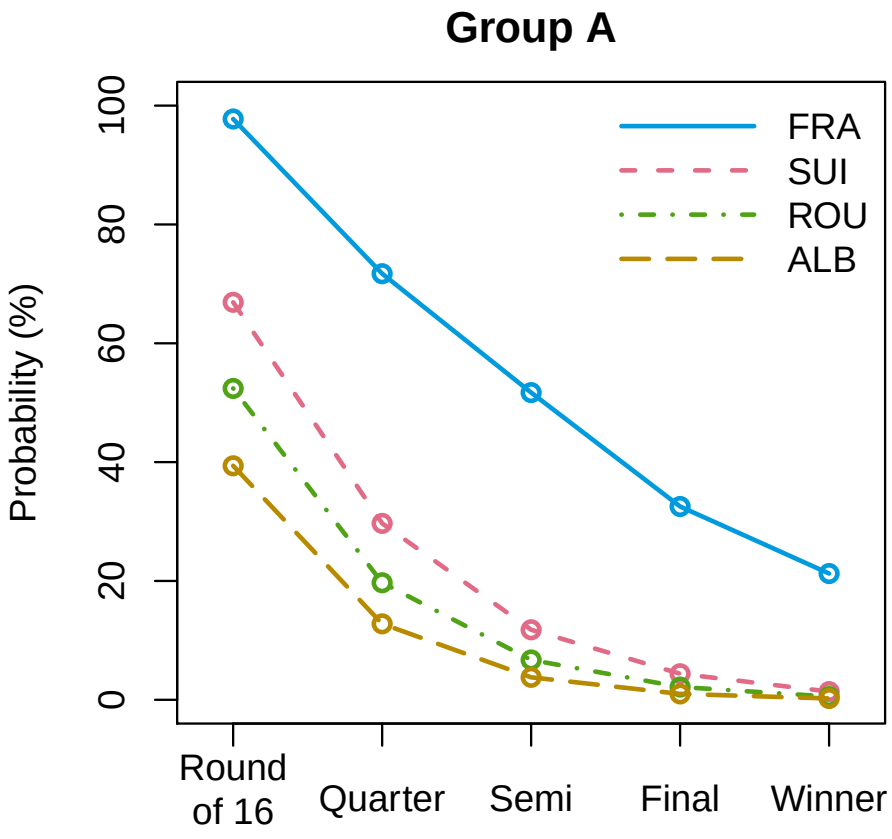
“Reverse” simulation:

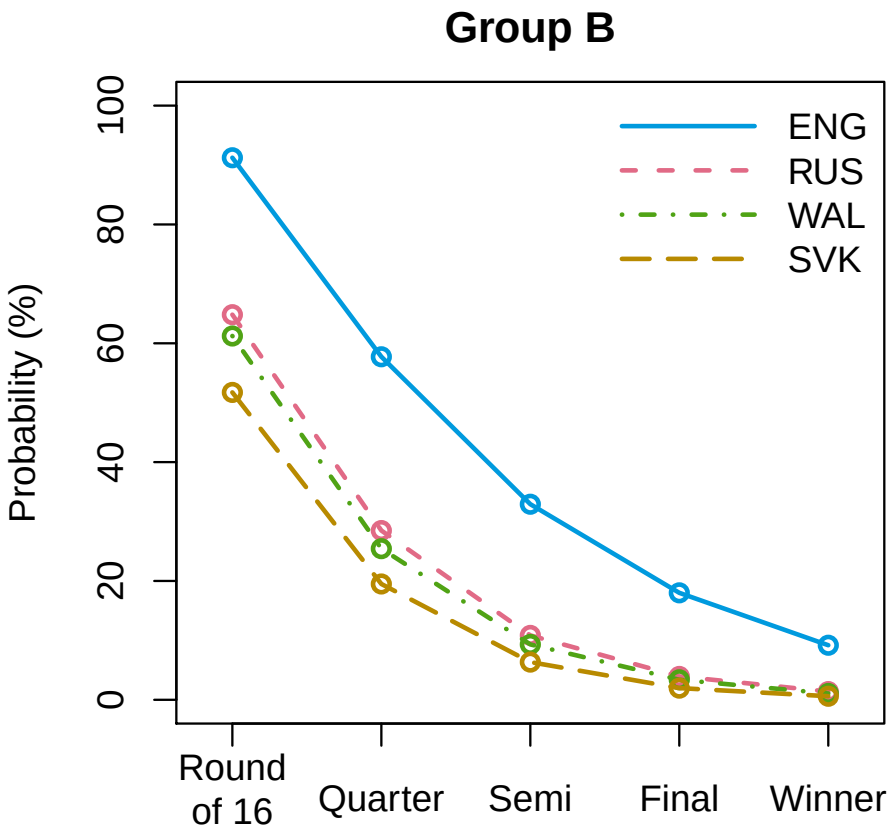
- If the team-specific *ability*$_i$ were known, pairwise probabilities $\pi_{i,j}$ could be computed.
- Given $\pi_{i,j}$ the whole tournament can be simulated (assuming abilities do not change and ignoring possible draws during the group stage).
- Using “many” simulations (here: 100,000) of the tournament, 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

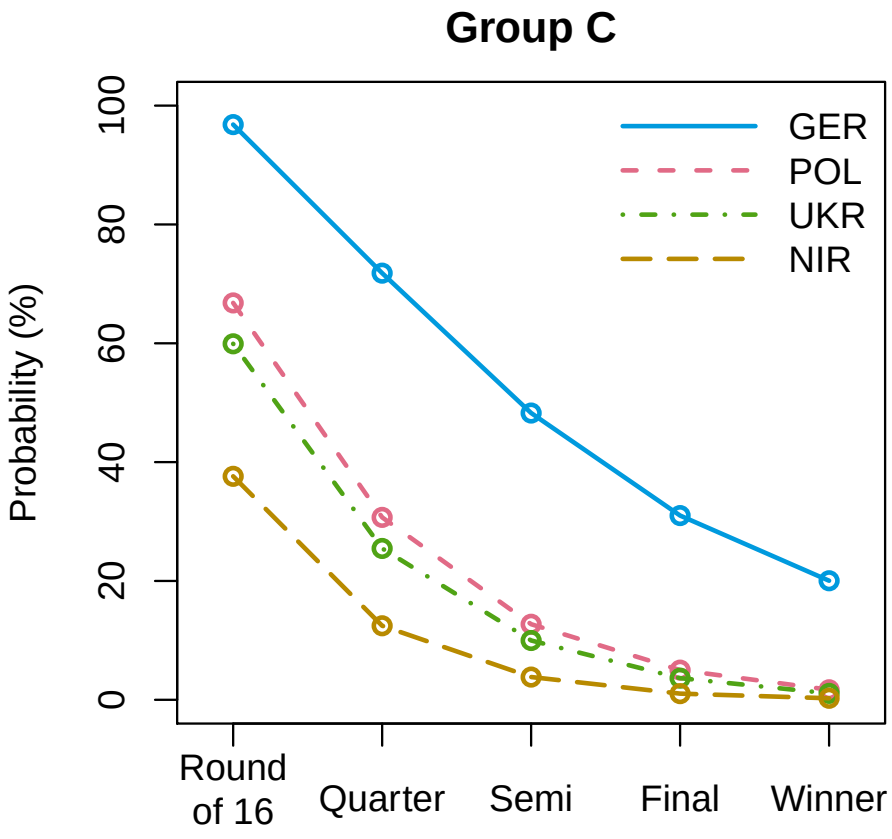


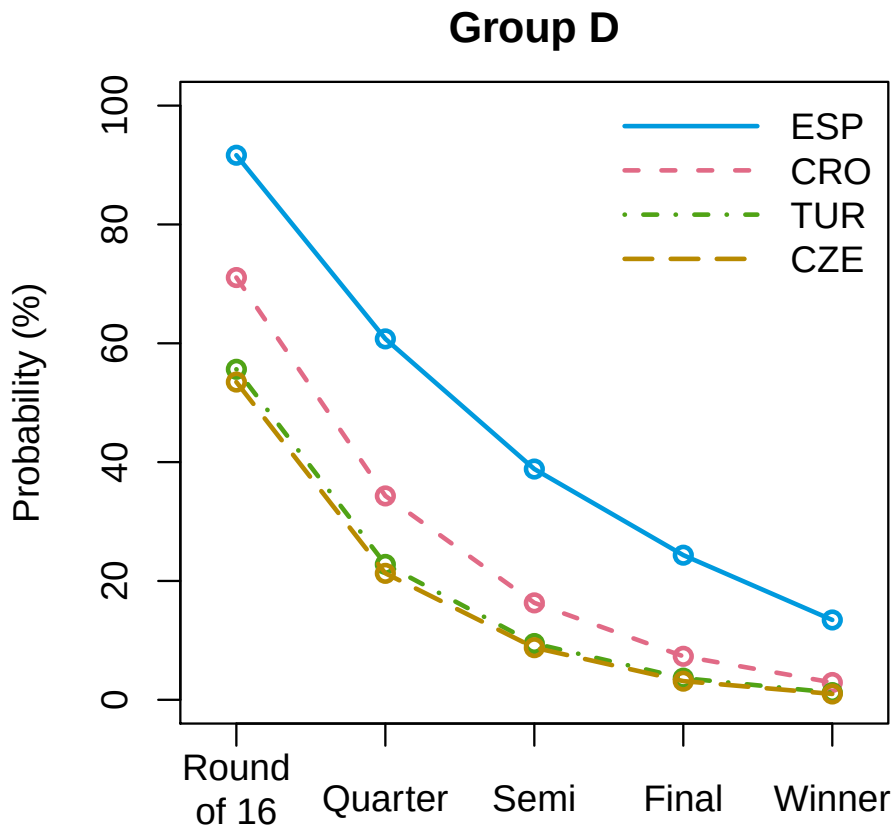
Tournament simulations: Survival curves



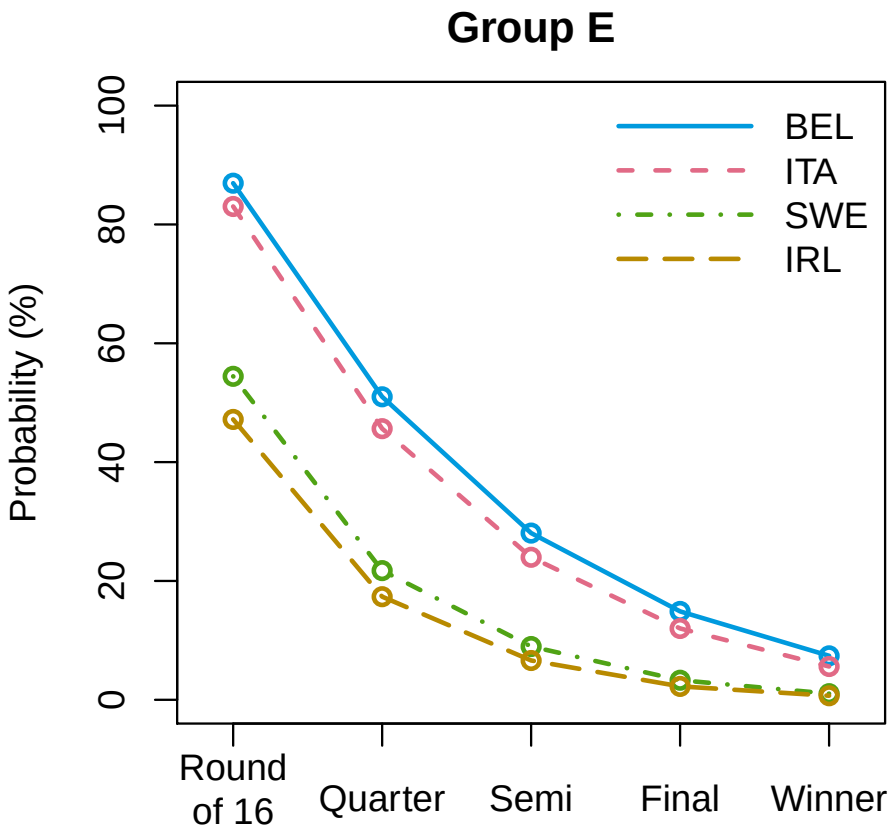


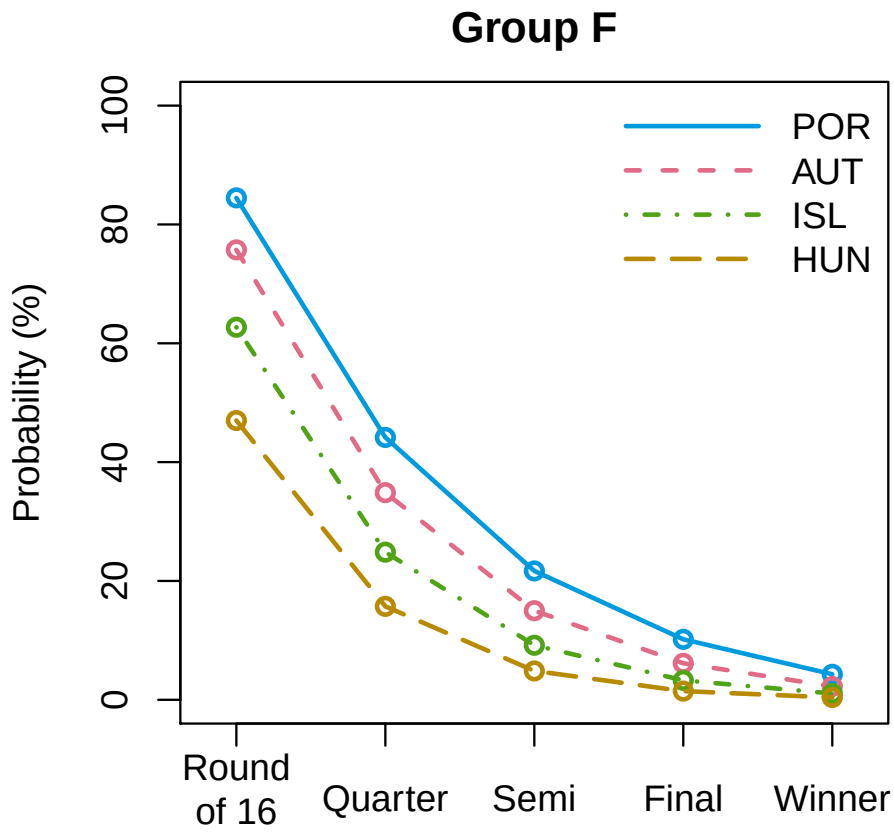
Tournament simulations: Survival curves





Tournament simulations: Survival curves





Outcome verification

Source: Spiegel.de

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:

$$ability_{Elo,i} = 10^{Elo_i/400}.$$

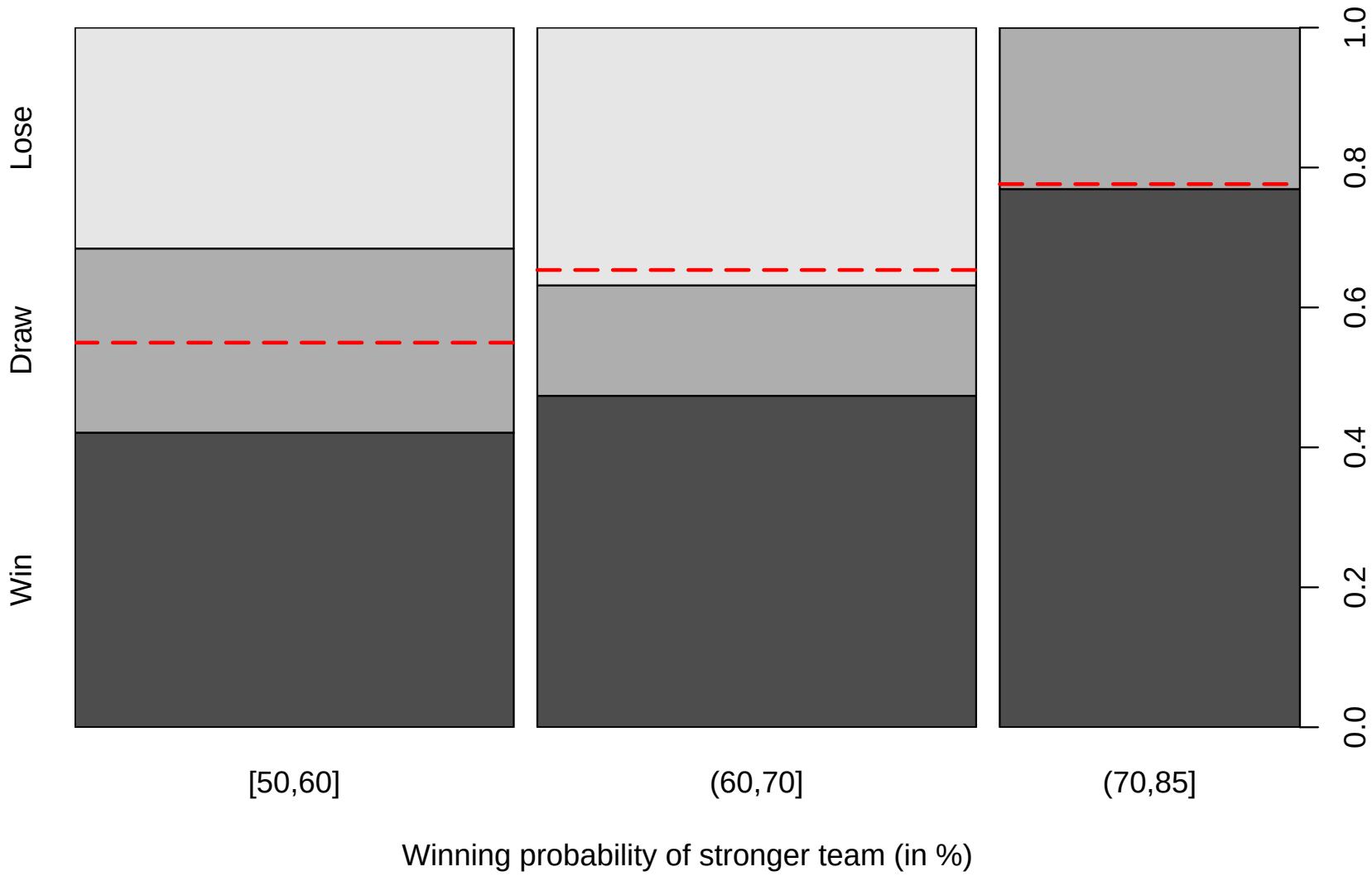
- Check whether pairwise probabilities roughly match empirical proportions from clusters of matches.

Outcome verification: Ranking

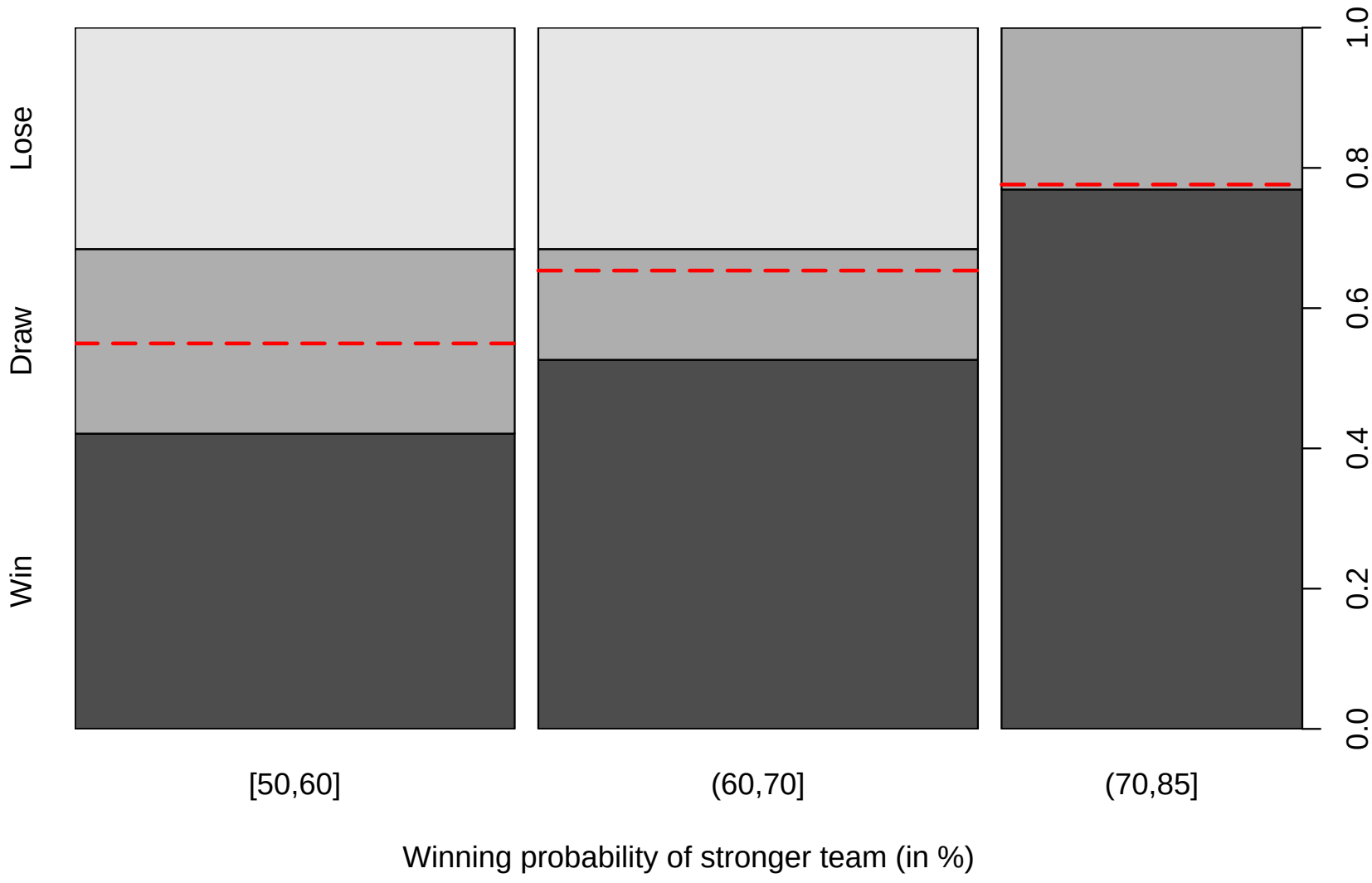
Spearman rank correlation of observed tournament ranking with bookmaker consensus model (BCM) as well as FIFA and Elo ranking:

BCM (Probabilities)	0.523
BCM (Abilities)	0.436
Elo (Probabilities)	0.344
Elo	0.339
FIFA	0.310

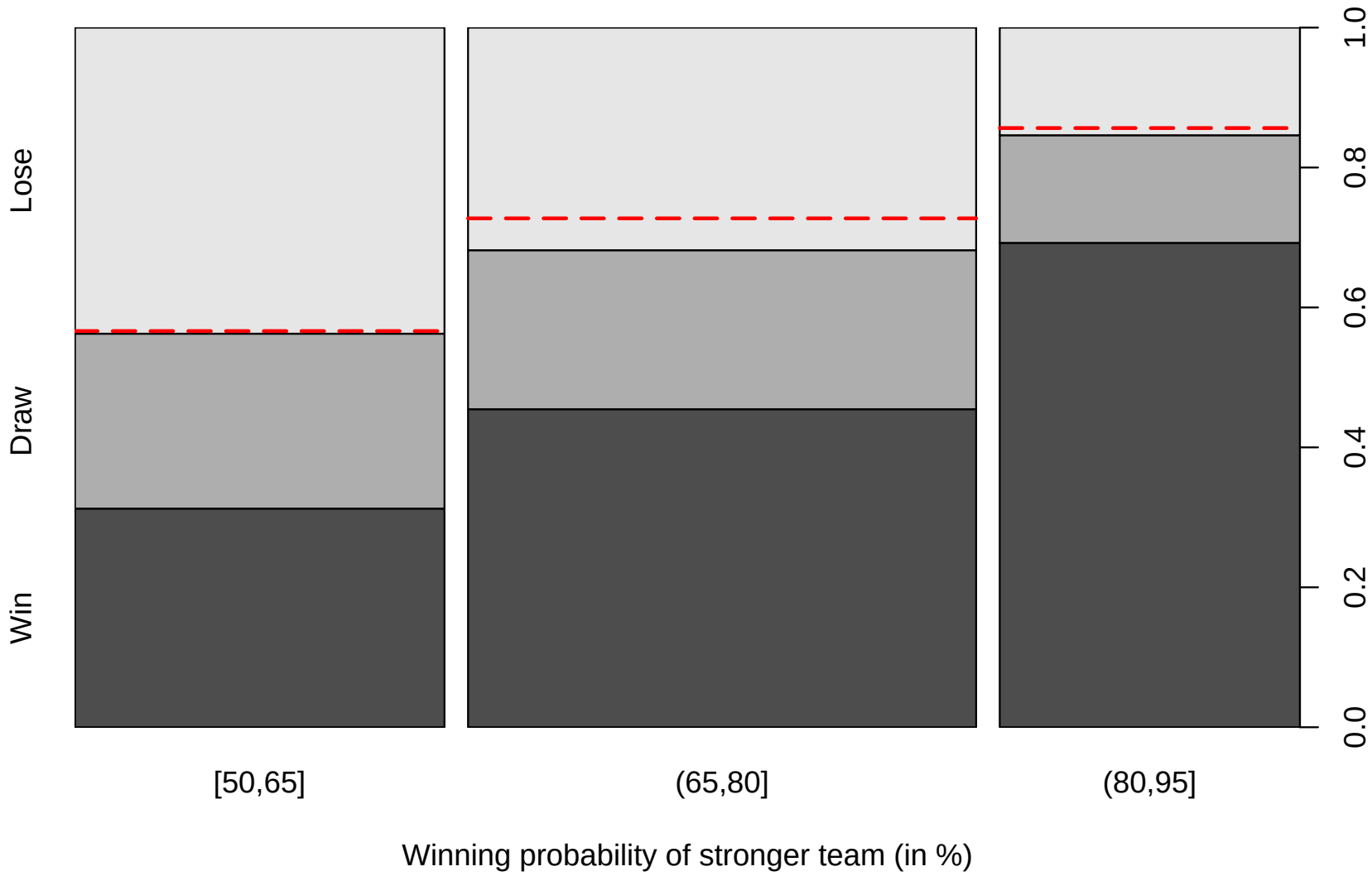
Outcome verification: BCM pairwise probabilities



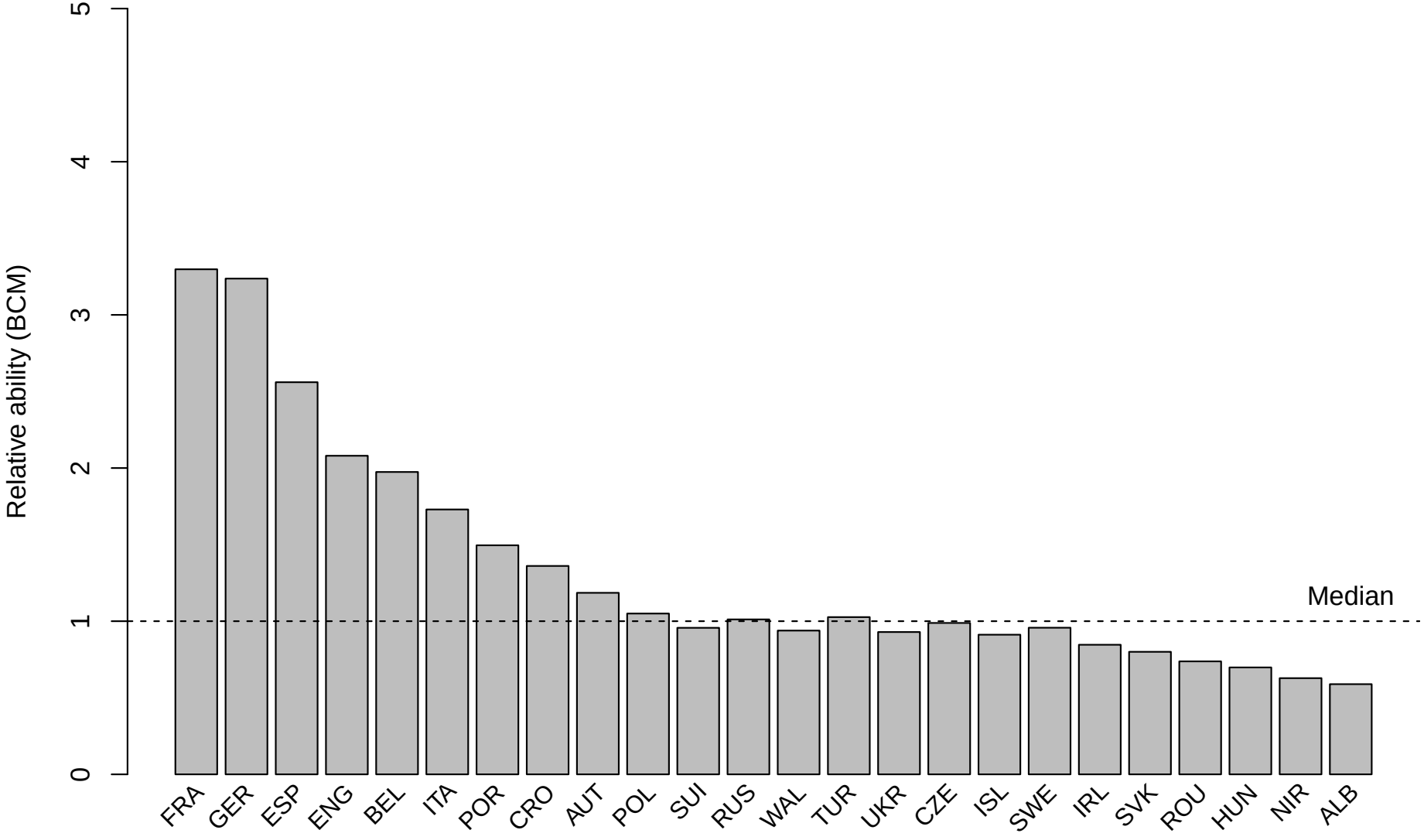
Outcome verification: BCM pairwise probabilities



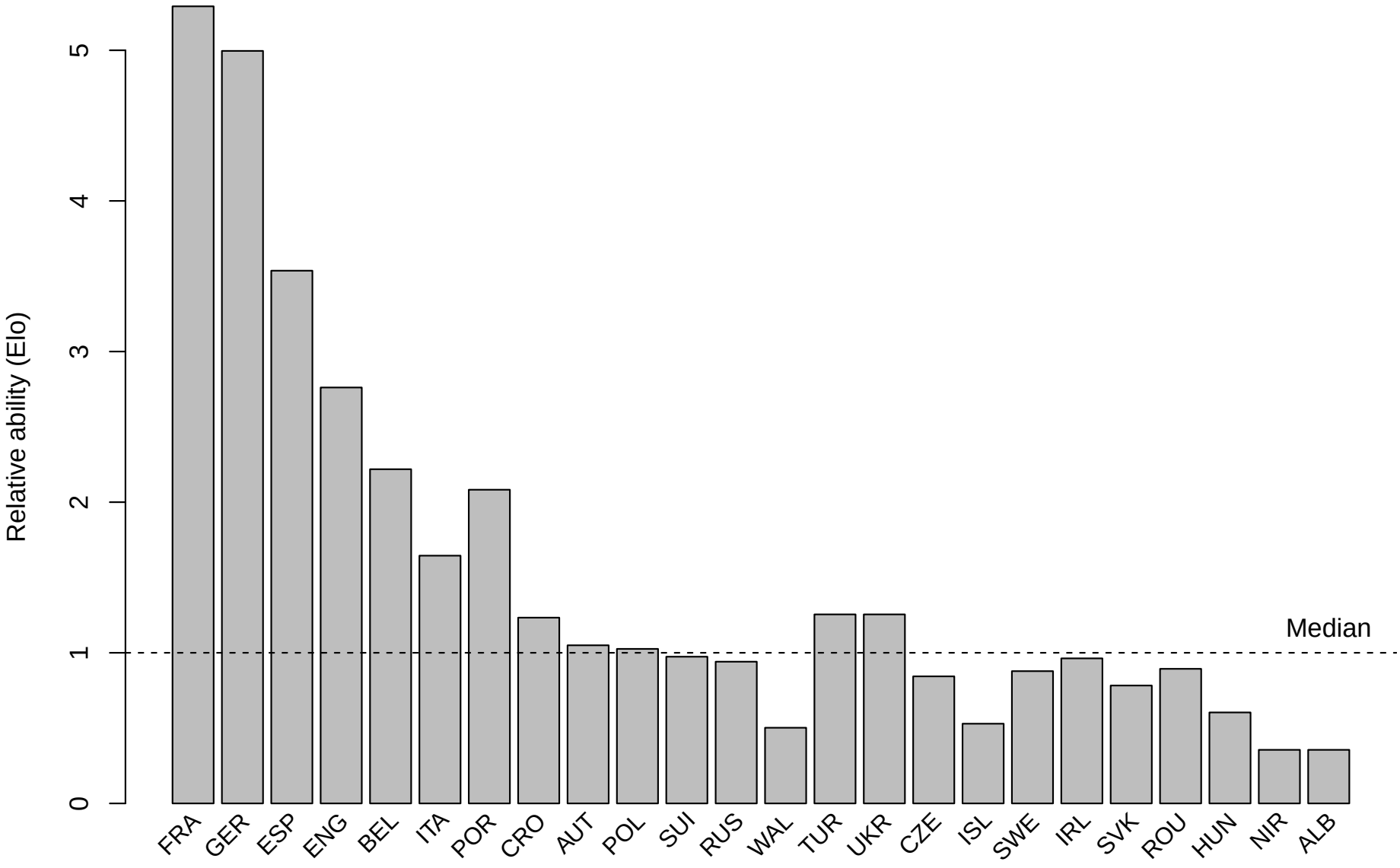
Outcome verification: Elo pairwise probabilities



Outcome verification: BCM abilities



Outcome verification: Elo abilities



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.

References

Zeileis A, Leitner C, Hornik K (2016). "Predictive Bookmaker Consensus Model for the UEFA Euro 2016." *Working Paper 2016-15*, Working Papers in Economics and Statistics, Research Platform Empirical and Experimental Economics, Universität Innsbruck. URL `http://EconPapers.RePEc.org/RePEc:inn:wpaper:2016-15`.

Leitner C, Zeileis A, Hornik K (2011). "Bookmaker Consensus and Agreement for the UEFA Champions League 2008/09." *IMA Journal of Management Mathematics*, **22**(2), 183–194. `doi:10.1093/imaman/dpq016`.

Leitner C, Zeileis A, Hornik K (2010). "Forecasting Sports Tournaments by Ratings of (Prob)abilities: A Comparison for the EURO 2008." *International Journal of Forecasting*, **26**(3), 471–481. `doi:10.1016/j.ijforecast.2009.10.001`.

Groups A and B

Rank	Team	Probability (in %)
1	**FRA**	**97.8**
2	**SUI**	**66.9**
3	ALB	39.4
4	ROU	52.4

Rank	Team	Probability (in %)
1	**WAL**	**61.2**
2	**ENG**	**91.2**
3	**SVK**	**51.7**
4	RUS	64.8

Groups C and D

Rank	Team	Probability (in %)
1	**GER**	**96.8**
2	**POL**	**66.8**
3	**NIR**	**37.6**
4	UKR	59.9

Rank	Team	Probability (in %)
1	**CRO**	**71.1**
2	**ESP**	**91.7**
3	TUR	55.6
4	CZE	53.5

Groups E and F

Rank	Team	Probability (in %)
1	**ITA**	**83.0**
2	**BEL**	**86.9**
3	**IRL**	**47.2**
4	SWE	54.4

Rank	Team	Probability (in %)
1	**HUN**	**47.0**
2	**ISL**	**62.7**
3	**POR**	**84.5**
4	AUT	75.7

Round of 16

Teams		Probability (in %)	Result
POL	SUI	50.6	6:5 (pen.)
WAL	NIR	61.1	1:0
POR	CRO	52.4	1:0 (a.e.t.)
FRA	IRL	79.6	2:1
GER	SVK	80.2	3:0
BEL	HUN	73.9	4:0
ESP	ITA	59.7	0:2
ENG	ISL	69.1	1:2

Quarterfinal, semifinal, final

Teams		Probability (in %)	Result
Quarterfinal			
POL	POR	41.2	4:6 (pen.)
WAL	BEL	33.4	3:1
GER	ITA	65.2	7:6 (pen.)
FRA	ISL	78.0	5:2
Semifinal			
POR	WAL	60.2	2:0
GER	FRA	49.5	0:2
Final			
POR	FRA	31.2	1:0 (a.e.t.)