

# EURO 2020 predictions: semi-finals

Leonardo Egidi - DEAMS, University of Trieste, Italy. Mail: legidi@units.it

4 July 2021

## Contents

<b>The statistical model (in brief)</b>	<b>1</b>
<b>Predictions: semi-finals (6-7 July)</b>	<b>2</b>
<b>Expected number of goals</b>	<b>2</b>
<b>Estimated attack/defence abilities</b>	<b>3</b>
<b>Euro 2020 winning probabilities</b>	<b>3</b>

## The statistical model (in brief)

We use a **double Poisson model with dynamic team-specific abilities** for the attack and the defence. Let  $(X_i, Y_i)$  denote the random number of goals scored by the home and the away team in the  $i$ -th game,  $i = 1, \dots, n$ , respectively. **ranking** denotes the Coca-Cola FIFA ranking at May 27th, 2021, whereas **att** and **def** denote the attack and the defence abilities, respectively.

$$X_i | \lambda_{1i} \sim \text{Poisson}(\lambda_{1i}), \quad (1)$$

$$Y_i | \lambda_{2i} \sim \text{Poisson}(\lambda_{2i}), \quad (2)$$

$$\log(\lambda_{1i}) = \text{home} + \text{att}_{h_i,t} + \text{def}_{a_i,t} + \frac{\gamma}{2}(\text{ranking}_{h_i} - \text{ranking}_{a_i}) \quad (3)$$

$$\log(\lambda_{2i}) = \text{att}_{a_i,t} + \text{def}_{h_i,t} - \frac{\gamma}{2}(\text{ranking}_{h_i} - \text{ranking}_{a_i}), \quad i = 1, \dots, n \text{ (matches)}, \quad (4)$$

$$\text{att}_{k,t} \sim \mathcal{N}(\text{att}_{k,t-1}, \sigma^2), \quad (5)$$

$$\text{def}_{k,t} \sim \mathcal{N}(\text{def}_{k,t-1}, \sigma^2), \quad (6)$$

$$\sum_{k=1}^{n_t} \text{att}_{k,t} = 0, \quad \sum_{k=1}^{n_t} \text{def}_{k,t} = 0, \quad k = 1, \dots, n_t \text{ (teams)}, \quad t = 1, \dots, T \text{ (times)}. \quad (7)$$

Lines (1)-(2) display the likelihood's equations (two Poisson distributions); lines (3)-(4) display the log-linear models for the scoring rates  $\lambda_1, \lambda_2$ ; lines (5)-(6) display the dynamic prior distributions for the attack and the defence parameters, respectively; line (7) displays the sum-to-zero identifiability constraints. Model fitting has been obtained through the Hamiltonian Monte Carlo sampling, 2000 iterations, 4 chains (**rstan** package). The historical data used to fit the models come from: **Nations' League** (2019-2020), **Euro UEFA Qualifiers** (2020-2021), **World Cup UEFA Qualifiers** (2021), **UEFA Euro 2020** (groupstage + round of 16 + quarter of finals matches).

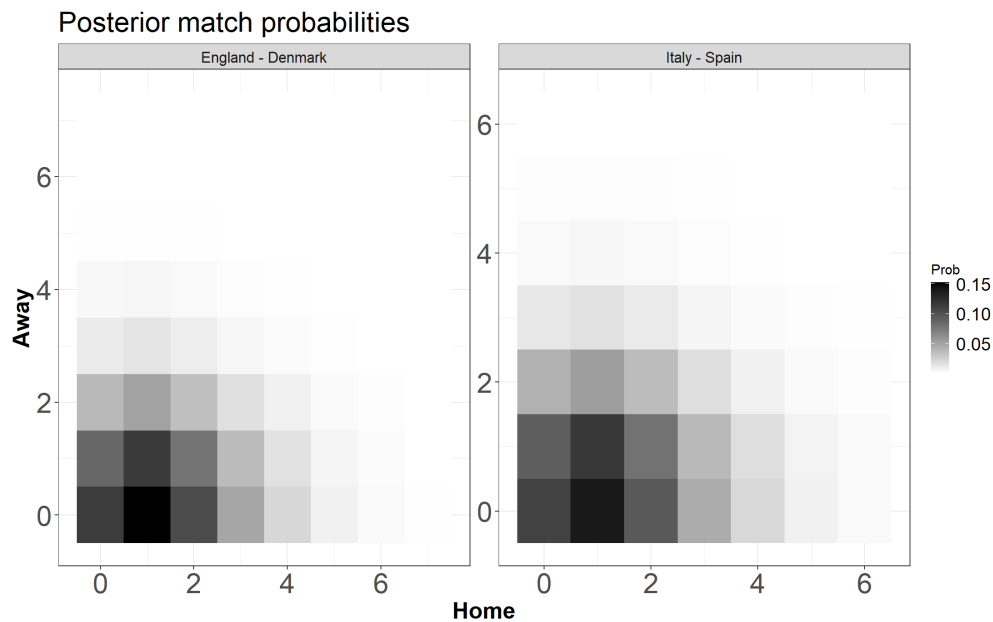
The idea is to provide a dynamic predictive scenario: at the end of each match-day, the model will be refitted to predict the remaining matches.

## Predictions: semi-finals (6-7 July)

Posterior matches probabilities from the posterior predictive distribution of the model above are displayed in the table below. **mlo** denotes the most likely exact outcome (in parenthesis, the corresponding posterior probability). Darker regions in the plots below denote more likely outcomes: on the  $x$ -axis the home goals, on the  $y$ -axis the away goals.

**Attention:** the matches probabilities below refer to the results within the regular 90 minutes.

home	away	home win	draw	away win	mlo
Italy	Spain	0.488	0.269	0.243	1-0 (0.139)
England	Denmark	0.505	0.269	0.226	1-0 (0.152)



## Expected number of goals

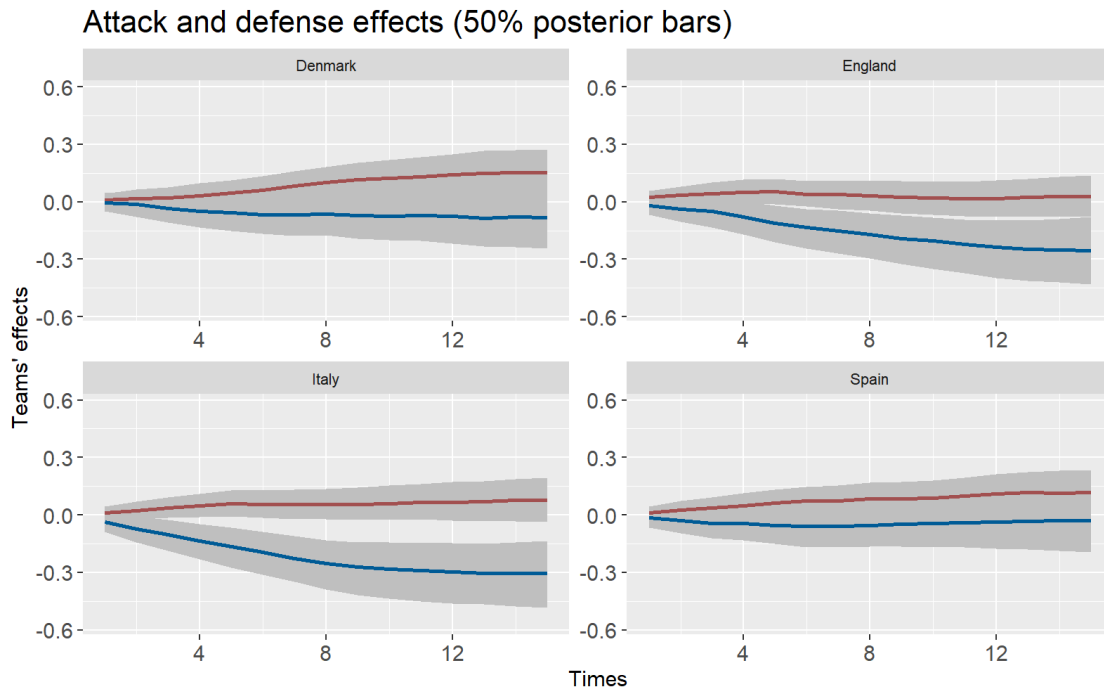
We compute also the **expected number of goals**  $\lambda_1, \lambda_2$  for each match, obtained by computing the median values from the MCMC sampling for the scoring rates.

**Attention:** these expected goals do not represent the most likely results according to posterior probabilities.

home	away	exp_home	exp_away
Italy	Spain	1.36	0.85
England	Denmark	1.41	0.81

## Estimated attack/defence abilities

In the plot below we display the posterior intervals for the **attack** (red) and **defence** (blue) abilities estimated through the training set matches, from **October 2019** until the **round of 16**: the higher the attack and the lower the defence values for a given team, and the better is the estimated overall team's ability.



## Euro 2020 winning probabilities

We compute the final winning probabilities for each team. We simulated in advance the two semi-finals and the final and we report the final winning probabilities expressed in percentages (%).

**Attention:** these probabilities come from an ahead-simulation of the semi-finals and the final based on 4000 MCMC iterations.

Winning team	Winning %
<b>Italy</b>	<b>34.2</b>
England	29.4
Spain	22.3
Denmark	14.0