

EURO 2024 knockout predictions: Semifinals

Leonardo Egidi - Giulio Fantuzzi | University of Trieste

09th July 2024

Contents

The statistical model (in brief)	1
Knockout predictions: Semifinals (09th July - 10th July)	2

The statistical model (in brief)

We use a **diagonal-inflated Bivariate-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 April 4th, 2024, whereas **att** and **def** denote the attack and the defence abilities, respectively.

$$(X_i, Y_i) \sim \begin{cases} (1-p)\text{BP}(x_i, y_i | \lambda_1, \lambda_2, \lambda_3) & \text{if } x \neq y \\ (1-p)\text{BP}(x_i, y_i | \lambda_1, \lambda_2, \lambda_3) + pD(x, \eta) & \text{if } x = y, \end{cases} \quad (1)$$

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

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

$$\log(\lambda_{3i}) = \rho, \quad (4)$$

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

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

$$\rho, \gamma \sim \mathcal{N}(0, 1) \quad (7)$$

$$p \sim \text{Uniform}(0, 1) \quad (8)$$

$$\sum_{k=1}^{n_t} \text{att}_k = 0, \quad \sum_{k=1}^{n_t} \text{def}_k = 0, \quad k = 1, \dots, n_t \text{ (teams)}, \quad t = 1, \dots, T \text{ (times)}. \quad (9)$$

Lines (1) displays the likelihood's equations (diagonal inflated bivariate Poisson); lines (2)-(4) display the log-linear models for the scoring rates λ_1, λ_2 and the covariance parameter λ_3 ; lines (5)-(6) display the dynamic prior distributions for the attack and the defence parameters, respectively; lines (7)-(8) display prior distributions for the other model parameters; line (9) displays the sum-to-zero identifiability constraints. Model fitting has been obtained through the Hamiltonian Monte Carlo sampling, 2000 iterations, 4 chains using the **footBayes** R package (with the underlying **rstan** package). The historical data used to fit the models come from *all the international matches played during the years' range 2020-2024*.

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.

Knockout predictions: Semifinals (09th July - 10th 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.

home	away	home win	draw	away win	mlo
Spain	France	0.333	0.308	0.359	0-0 (0.154)
Netherlands	England	0.244	0.262	0.495	0-1 (0.127)

