EURO 2024 groupstage predictions: 2nd match day

Leonardo Egidi - Giulio Fantuzzi | University of Trieste

18th June 2024

Contents

The statistical model (in brief)

1

Groupstage predictions: 2nd match day (19-22 June)

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, $i = 1, \ldots, n$, respectively. ranking denotes the Coca-Cola FIFA ranking at April 4th, 2024, whereas att and def denote the attack and the defence abilities, respectively.

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

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

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

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

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

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

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

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

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

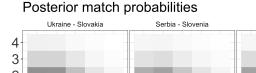
Lines (1) displays the likelihood's equations (diagonal inflated bivariate Poisson); lines (2)-(4) display the log-linear models for the scoring rates λ_1, λ_2 and the covariance parameter λ_3 ; lines (5)-(6) display the dynamic prior distributions for the attack and the defence parameters, respectively; 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.

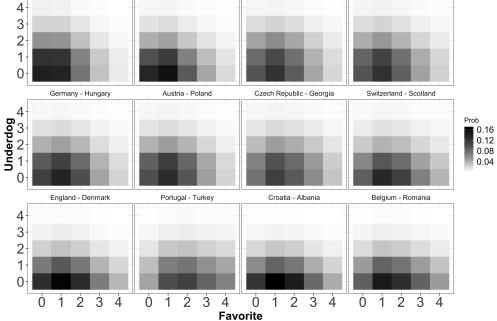
The idea is to provide a dynamic predictive scenario: at the end of each match-day, the model will be refitted to predict the remaining matches.

Groupstage predictions: 2nd match day (19-22 June)

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
Croatia	Albania	0.624	0.243	0.132	1-0 (0.17)
Germany	Hungary	0.482	0.274	0.244	1-0 (0.137)
Scotland	Switzerland	0.187	0.239	0.574	$0-1 \ (0.133)$
Slovenia	Serbia	0.248	0.292	0.460	0-1 (0.158)
Denmark	England	0.157	0.260	0.583	0-1 (0.169)
Spain	Italy	0.475	0.264	0.261	1-0 (0.124)
Slovakia	Ukraine	0.314	0.304	0.382	0-0(0.14)
Poland	Austria	0.244	0.268	0.489	0-1 (0.129)
Netherlands	France	0.269	0.264	0.467	0-1 (0.123)
Georgia	Czech Republic	0.227	0.242	0.530	0-1 (0.112)
Turkey	Portugal	0.118	0.183	0.699	0-2(0.11)
Belgium	Romania	0.649	0.221	0.130	1-0 (0.147)





Spain - Italy