EURO 2024 groupstage predictions: 1st match day

Leonardo Egidi - Giulio Fantuzzi | University of Trieste

12th June 2022

Contents

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. $\mathsf{ranking}$ denotes the Coca-Cola FIFA ranking at October 6th, 2022, whereas att and def denote the attack and the defence abilities, respectively.

$$
\begin{aligned}
(X_i, Y_i) \ &\sim \begin{cases} (1-p)\mathrm{BP}(x_i, y_i|\lambda_1, \lambda_2, \lambda_3) & \text{if } x \neq y \\ (1-p)\mathrm{BP}(x_i, y_i|\lambda_1, \lambda_2, \lambda_3) + pD(x, \eta) & \text{if } x = y, \end{cases}, && (1)\\
\log(\lambda_{1i}) =&\ \mathrm{att}_{h_i,t} + \mathrm{def}_{a_i,t} + \frac{\gamma}{2}(\mathsf{ranking}_{h_i} - \mathsf{ranking}_{a_i}) && (2)\\
\log(\lambda_{2i}) =&\ \mathrm{att}_{a_i,t} + \mathrm{def}_{h_i,t} - \frac{\gamma}{2}(\mathsf{ranking}_{h_i} - \mathsf{ranking}_{a_i}), \quad i = 1, \dots, n \text{ (matches)}, && (3)\\
\log(\lambda_{3i}) =&\ \rho, && (4)\\
\mathrm{att}_{k,t} \sim&\ \mathcal{N}(\mathrm{att}_{k,t-1}, \sigma^2), && (5)\\
\mathrm{def}_{k,t} \sim&\ \mathcal{N}(\mathrm{def}_{k,t-1}, \sigma^2), && (6)\\
\rho,\ \gamma \sim&\ \mathcal{N}(0, 1) && (7)\\
p \sim&\ \mathrm{Uniform}(0, 1) && (8)\\
\sum_{k=1}^{n_t} \mathrm{att}_{k,} = 0,&\ \sum_{k=1}^{n_t} \mathrm{def}_{k,} = 0, \quad k = 1, \dots n_t \text{ (teams)},\ t = 1, \dots, T \text{ (times)}. && (9)
\end{aligned}
$$

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.

home	away	home win	draw	away win	mlo
Germany	Scotland	0.582	0.242	0.177	1-0 (0.142)
Hungary	Switzerland	0.327	0.322	0.351	0-0 (0.173)
Spain	Croatia	0.468	0.284	0.248	1-0 (0.15)
Italy	Albania	0.725	0.190	0.085	2-0 (0.149)
Poland	Netherlands	0.156	0.210	0.635	0-2 (0.113)
Slovenia	Denmark	0.181	0.266	0.553	0-1 (0.166)
Serbia	England	0.112	0.210	0.678	0-1 (0.146)
Romania	Ukraine	0.263	0.277	0.460	0-1 (0.132)
Belgium	Slovakia	0.731	0.188	0.081	2-0 (0.158)
Austria	France	0.169	0.240	0.591	0-1 (0.139)
Turkey	Georgia	0.484	0.239	0.276	1-1 (0.096)
Portugal	Czech Republic	0.691	0.194	0.115	1-0 (0.122)