Predictions for the EUROPEAN FOOTBALL CHAMPIONSHIP 2024 based on Statistical Analytical Football Models

AUEB & Trieste Sports Analytics Research Group,

Athens University of Economics and Business and

University of Trieste

This article was edited and co-authored by Ioannis Ntzoufras, Professor of Statistics at AUEB, and Argyro Damoulaki, PhD Candidate in the same department. The article is based on the analysis of the collaborating team of Trieste (Professors Leonardo Egidi and Nicola Torelli, PhD candidate Roberto Macrì Demartino, and data science master student Giulio Fantuzzi) with the assistance of V. Palaskas (OpenBet, application development) D. Karlis (AUEB Statistics, analysis consultant). The final result is a cooperation between the research teams of the two universities on Sports Analytics.

***

The first matchday of the European Championship was delightfully complete: some easy victories (Germany, Switzerland, Spain and Romania), reversals of results (Portugal prevailed in the Czech Republic), surprises (with Slovakia beating Belgium), beautiful goals (Turkey scored impressively against Georgia) and some favourites who managed to prevail "*just by getting the job done*" (Italy, Netherlands, England and France). Looking forward to the future, we will make a brief review of our predictions for the first matchday and present our "predictions" for the second matchday of the competition.

Reminder for friends of Statistics

The use of statistical techniques to predict football matches first appeared in the scientific literature in 1968 with the pioneering scientific publication of Reep & Benjamin. The next real innovations appear in the 80s (with the work of Michael Maher) and the 90s (with the work of Lee in 1997). However, the first important publications in the field, introducing models on which models are based and which we still use today, were the works of Dixon & Coles in 1997 and the bivariate Poisson model of Karlis and Ntzoufra in 2003 (two of the authors of this analysis). These two models formed the basis of modern models for predicting football match outcomes.

In this analysis we use the model of Karlis and Ntzoufras through the package "footbayes" in the statistical programming language R developed by Professor Leonardo Egidi from the University of Trieste with the assistance of Vasilis Palaskas (Analyst at Open Bet and active member of AUEB Sports Analytics Group). The model also includes the estimation of parameters that estimate the performance of each group that change over time. To learn the

model, all international matches of the 2020-2024 period were used. The main explanatory variable is the difference between the two teams in the Coca-Cola/FIFA ranking. The model, first proposed by Karlis & Ntzoufras in 2003, extends the usual two-variate Poisson model. Details of the statistical and machine learning model used can be found at the end of this article.

Matchday 1 Report

The predictions of the first 12 matches (matchday 1) and the final results are provided in Table 1. The model managed to correctly predict (based on the probability of possible outcomes of the match i.e. Win, draw, defeat) an important number of matches (75%) among which we single out the match between Hungary and Switzerland. This match was the closest according to our model, giving a slight lead to Switzerland who eventually prevailed. Also, in 2 more races the model showed the final result to be quite likely (but not the most likely). More specifically, in the Slovenia-Denmark match we have a remarkable chance of a draw (27%), while in the Romania-Ukraine match the probability of winning was remarkable for Ukraine (47%) but also high for Romania (25%). As for Slovakia's big surprise, the model gave only an 8% chance of winning just like any reasonable model we could make based on data. It should be noted here that a logical model of statistical and machine learning will in no way be able to catch surprises such as those that may occur due to randomness or specific situations that are not considered by the model and the data on which they have been trained.

Table 1: Table with the odds of the outcome of the matches for Matchday 1 of the European Championship 2024.

Rival teams (A-B)		Odds			Prevalent Result (Probability)	Final Result
		Win A Group	Draw	Niki B Group		
Germany	Scotland	0.579	0.243	0.178	1-0 (0.143)	5-1
Hungary	Switzerlan	0.326	0.329	0.345	0-0 (0.176)	1-3
Spain	Croatia	0.455	0.289	0.256	1-0 (0.140)	3-0
Italy	Albania	0.723	0.190	0.088	2-0 (0.148)	2-1
Poland	Netherlan	0.156	0.214	0.630	0-2 (0.113)	1-2
Slovenia	Denmark	0.186	0.270	0.543	0-1 (0.167)	1-1
Serbia	England	0.107	0.212	0.681	0-1 (0.150)	0-1
Romania	Ukraine	0.254	0.277	0.469	0-1 (0.137)	3-0
Belgium	Slovakia	0.729	0.190	0.081	2-0 (0.158)	0-1
Austria	France	0.170	0.243	0.588	0-1 (0.145)	0-1
Turkey	Georgia	0.491	0.240	0.269	1-0 (0.097)	3-1
Portugal	Czech Republic	0.693	0.196	0.111	2-0 (0.134)	2-1

Matchday 2 predictions

We proceed with optimism for the second matchday, therefore, with the model's predictions presented in Table 2.

From this table we distinguish the race

- Slovakia - Ukraine

as the most ambivalent.

As favorites stand out

1. Portugal with a 69% chance of winning against Turkey
2. Belgium with a 65% chance of winning against Romania

3. Croatia with a 62% chance of winning against Albania
4. England with a 58% chance of winning against Denmark
5. Switzerland with a 57% chance of winning against Scotland
6. Czech Republic with a 53% chance of winning against Georgia

Finally, we have five more races that are relatively close but with a slight lead of one of the two teams. In these matches we consider that the teams are relatively close and may even draw due to tactics and strategy. In particular, we have

1. Austria (49%) prevailing over Poland (24%)
2. Germany (48%) beating Hungary (24%)
3. Spain (47.5%) prevailing over Italy (26%)
4. France (47%) prevailing over the Netherlands (27%)
5. Serbia (46%) prevailing over Slovenia (45%)

Table 2: Table with the odds of the outcome of the matches for Matchday 2 of the European Championship 2024.

Rival teams (A-B)		Odds			Prevalent
		Win A Group	Draw	Niki B Group	Result (Probability)
Croatia	Albania	0.624	0.243	0.132	1-0 (0.170)
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 (0.140)
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	0.227	0.242	0.530	0-1 (0.112)
Turkey	Portugal	0.118	0.183	0.699	0-2 (0.110)
Belgium	Romania	0.649	0.221	0.130	1-0 (0.147)

Chart 1 gives in more detail the odds for each score for each of the 12 matches of Matchday 2.

Πιθανότητα αποτελεσμάτων αγώνα (αφού λάβουμε υπόψη μας τα δεδομένα)
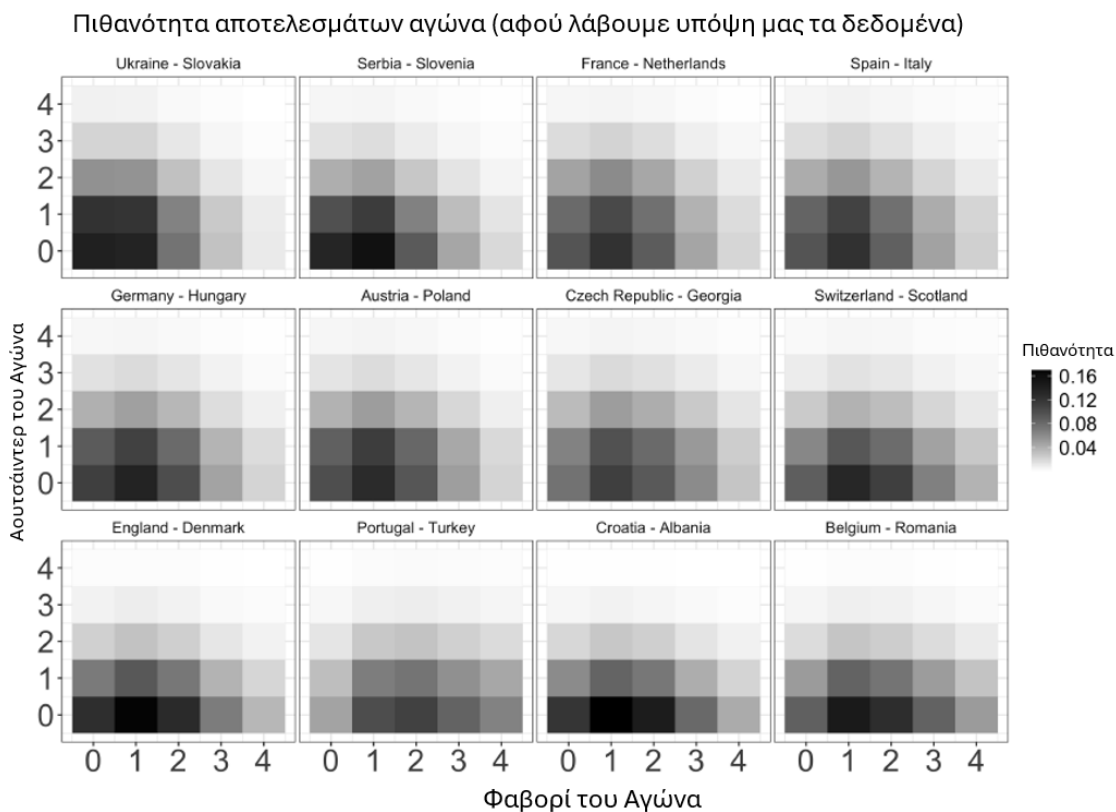
Diagram 1: Probability Chart of possible scores for the Games of Matchday 2 of the European Championship 2024.

Bibliography for reading fans

- Dixon, M.J. and Coles, S.G. (1997), Modelling Association Football Scores and Inefficiencies in the Football Betting Market. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, **46**, 265-280.
- Karlis, D. and Ntzoufras, I. (2003), Analysis of sports data by using bivariate Poisson models. *Journal of the Royal Statistical Society: Series D (The Statistician)*, **52**, 381-393.
- Lee A.J. (1997). Modeling Scores in the Premier League: Is Manchester United Really the Best? *Chance*, **10**, 15-19.
- Maher, M.J. (1982), Modelling association football scores. *Statistica Neerlandica*, **36**, 109-118.
- Reep, C., & Benjamin, B. (1968). Skill and Chance in Association Football. *Journal of the Royal Statistical Society. Series A (General)*, **131**, 581-585.

The Magic Equations of the statistical model

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

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

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

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

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

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

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

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

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

- i is the indicator of the struggle
- X_i and is the number of goals scored between the 1st Y_i and 2nd teams in match i
- h_i and are the 1st a_i and 2nd teams respectively (or home and away teams – where applicable) for the i match.
- $\mathrm{att}_{k,t}$ and parameters estimating the offensive and defensive capacity/capability of group k at time $\mathrm{def}_{k,t}$ t (dynamic parameters that change over time)
- $\mathrm{ranking}_k$ Coca-Cola FIFA ranking on April 4, 2024 for team k.

A few words about the Authors

AUEB Sports Analytics Group

Ioannis Ntzoufras is Professor of Statistics and Chairman of the Department of Statistics of the Athens University of Economics and Business. He is a founding member of AUEB Sports Analytics Group along with Dimitris Karlis. He has recognized scientific activity in areas such as Bayesian statistical methodology, computational statistics, biostatistics, psychometrics and sports analytics.

Dimitris Karlis is Professor of Statistics and Deputy Chairman of the Department of Statistics of the Athens University of Economics and Business. He is a founding member of AUEB Sports Analytics Group along with Ioannis Ntzoufras. He has recognized scientific activity in areas such as statistical methodology, computational statistics, biostatistics, and sports analytics.

Argyro Damoulaki is a PhD candidate at the Department of Statistics of AUEB. Her current research focuses on performance evaluation models using Bayesian statistics. He is involved in sports analytics and has been an active member of AUEB Sports Analytics Group since 2022.

Vasilis Palaskas is a Statistical Analyst and Data Scientist at Open Bet. He previously worked as head of analysis and forecasting at Fantasy Sports Interactive (FSI). He has been an active member of the AUEB Sports Analytics research team since 2019 when he graduated from M.Sc. in Statistics of the Athens University of Economics and Business.

The research team of the University of Trieste

Leonardo Egidi is Assistant Professor of Statistics at the University of Trieste in Italy and a member of the AUEB Sports Analytics Group of the Athens University of Economics and Business. He has a PhD in football modeling and analytics and intense research activity in Bayesian Statistical methodology.

Nicola Torelli is Full Professor of Statistics at the University of Trieste. His main research fields of statistical classification, clustering and Bayesian modeling. He is the former president of the Italian Statistical Society.

Roberto Macri Demartino is a PhD candidate at the University of Padua. His main research interests focus on Bayesian statistical methodology with applications in a variety of fields, such as sports analytics.

Giulio Fantuzzi received a degree in Statistics and is now a postgraduate student in Data Science and Artificial Intelligence at the University of Trieste. He is very active in the field of sports modeling mainly using R and Python.

Active Author Collaborations

The three authors (L. Egidi, I. Ntzoufras and D. Karlis) of the article are currently collaborating on the writing of a book on *Football Analytics* in an international scientific house while in the last workshop of the team they gave a seminar course in Football analytics.

The two universities are also preparing a postgraduate program in Sports Analytics with plans to start in October 2025.

L. Egidi and **V. Palaskas** collaborate on the development of the software "footbayes" (library of the statistical programming language R).

 L. Egidi, I. Ntzoufras and **V. Palaskas** collaborate on the writing of a scientific article evaluating players in Volleyball.

V. Palaskas is a graduate of AUEB M.Sc. in Sports Analytics, an active member of AUEB Sports Analytics and has co-supervised a diploma thesis in the framework of AUEB's M.Sc. in Statistics and his previous collaboration with FSI (Fantasy Sports Interactive).

Η Ομάδα AUEB Sports Analytics

AUEB **Sports Analytics Group** was founded in 2015 by professors Ioannis Ntzoufras and Dimitris Karlis. Its members are important members of the sports analytics community such as Leonardo Egidi (University of Trieste), Ioannis Kosmidis (Warwick), Konstantinos Pelechrinis (Pittsburg), Nial Friel (UCD) and Gianluca Baio (UCL) as well as the former coach of the Greek national volleyball team. Sotiris Drikos and Sacramento Kings scouter Christos Marmarinos. The research team is responsible for the series of annual conferences named AUEB Sports Analytics Workshop (6 in total) and in 2019 organized the international conference MathSport 2019 with 200 participating scientists from around the world. The team has a number of important scientific publications in the field of sports analytics. Finally, we would like to mention that the team was founded in 2015 due to the visit of Professor Stefan Kesenne (University of Antwerp & Leuven), a great Sports Economist who played an active role in the Bosman case. Stefan Kesenne actively supported the team until 2021 when he suddenly passed away. The existence of AUEB Sports Analytics Group is largely due to the contribution and inspiration that Mr. Kesenne gave us.

Website του group https://aueb-analytics.wixsite.com/sports

Website του AUEB Sports Analytics Workshop https://aueb-analytics.wixsite.com/saw2024

Listen to the podcast on Basketball Analytics by Ioannis Ntzoufras, Argyro Damoulakis and Christos Marmarinos here.