



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 (Professor Leonardo Egidi and PhD candidates Roberto Macri Demartino and Giulio Fantuzzi) with the assistance of V. Palaskas (OpenBet, application development) D. Karlis (AUEB Statistics, analysis consultant). The final result is cooperation between the research teams of the two universities on Sports Analytics.

The quarter-finals of the European Championship 2024 are here and great matches await us. Two classic "derbies", between Spain-Germany and Portugal-France, and two teams ready to qualify, Switzerland and Turkey, against the traditional powers of England and the Netherlands. After a brief review of our predictions for the round of 16, we present our predictions for the quarter-finals.

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.

Review of the round of 16.

With the completion of the round of 16, the results are quite satisfactory, since the model managed to correctly predict 75% of the games. The games were basically close, like all knockout matches, while the favorites did not prevail so easily. France won Belgium with an own goal in the 85th minute, England tied in stoppage time and won Slovakia in extra time while Portugal needed a penalty shootout to break Slovenia's resistance. Spain, Germany and the Netherlands easily prevailed against their opponents. The big surprise of the round of 16 was Switzerland, who comfortably prevailed against Italy. Turkey's surprise-victory against Austria was similar, but it seemed difficult to overcome until the end. It should be noted that the win probability of Switzerland based on the model was remarkable (29%), as well as Turkey's that was not negligible (22%).

Rival teams (A-B)	Odds			Prevalent Result (Probability)	Final Result
	Win A Group	Draw	Niki B Group		
Switzerland Italy	0.288	0.273	0.439	0-1 (0.123)	2 – 0
Germany Denmark	0.448	0.263	0.289	1-0 (0.120)	2 – 0
England Slovakia	0.714	0.206	0.080	1-0 (0.160)	2 – 1
Spain Georgia	0.726	0.186	0.088	2-0 (0.139)	4 – 1
France Belgium	0.406	0.301	0.293	0-0 (0.152)	1 – 0
Portugal Slovenia	0.653	0.220	0.127	1-0 (0.145)	0 – 0
Romania Netherland	0.163	0.213	0.624	0-1 (0.109)	0 – 3
Austria Turkey	0.550	0.231	0.218	1-0 (0.101)	1 – 2

Table 1: Table with the odds of the outcome of the matches for the round of 16 of the European Championship 2024.

Predictions for the Round of 8

Heading into the finals, the differences between the teams are smaller and the matches more close. From Table 2 with the possible results, the following teams stand out as favorites:

1. The Netherlands with a 61% chance of winning against Turkey
2. England with a 53% chance of winning against Switzerland

Of these two favorites, the Netherlands had shown a mediocre face in the group stage but seem to have turned the corner in the knockouts. England, on the other hand, troubled in the round of 16 at a time when Switzerland looks very strong as an opponent, and it seems based on the matches that they will make it much more difficult for England than the model predicts. In fact, the probability of a draw is increased (28%).

Finally, the remaining two matches are more close but with a slight lead of one of the two teams. In these races we think the teams are relatively close. Specifically, we have:

1. Spain (45%) prevailing over Germany (27%)
2. France (45%) prevailing over Portugal (27%)

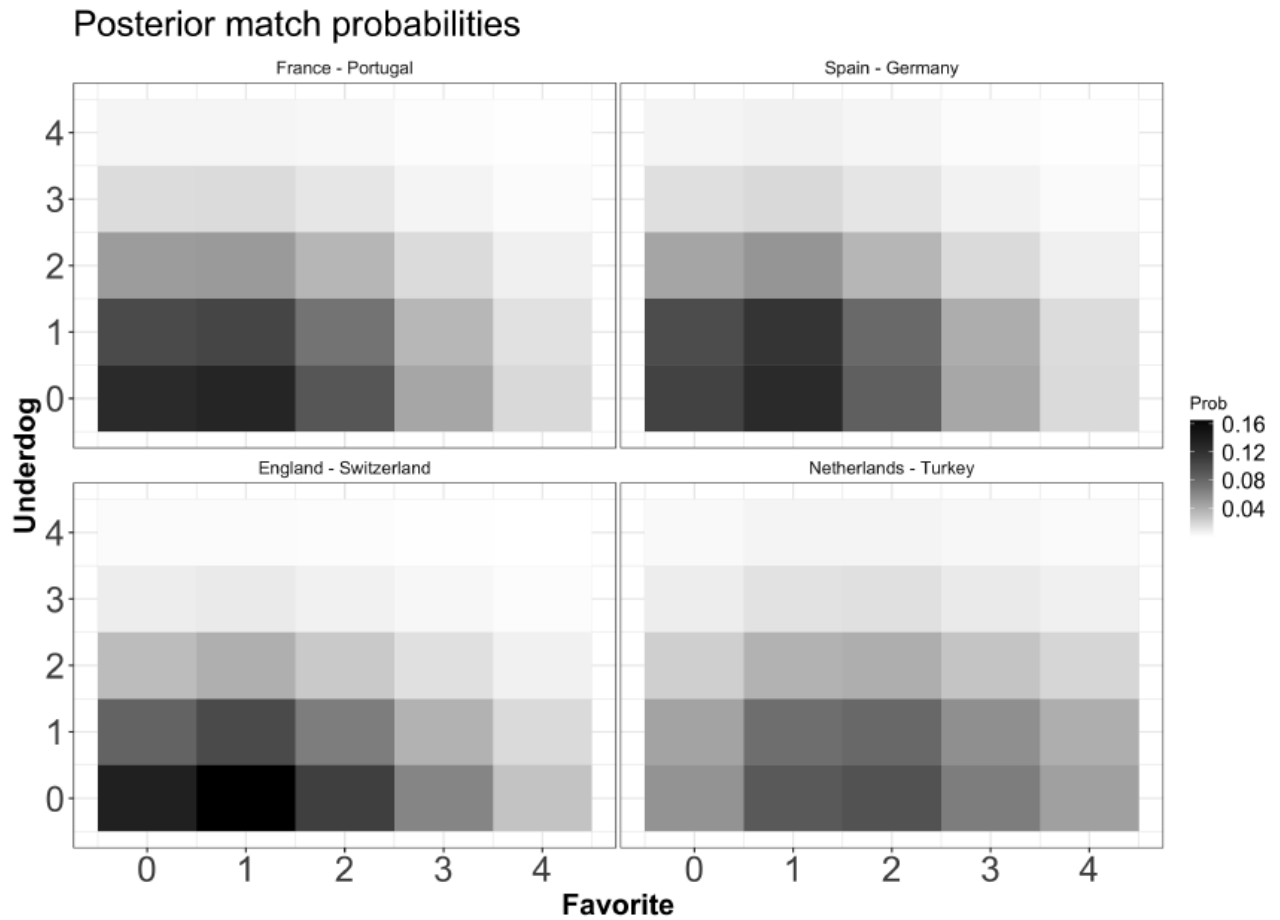
For these two matches, the probability of a draw is increased (28%) and any outcome is not unlikely.

Table 2: Table with the odds of the outcome of the matches for the round of 8 of the European Championship 2024.

Rival teams (A-B)		Odds			Prevalent Result (Probability)
		Win A Group	Draw	Niki B Group	
Spain	Germany	0.453	0.276	0.271	1-0 (0.127)
Portugal	France	0.270	0.283	0.447	0-1 (0.131)
England	Switzerland	0.530	0.277	0.193	1-0 (0.165)
Netherlands	Turkey	0.606	0.210	0.184	2-0 (0.095)

Figure 1 gives in more detail the odds for each score for each of the 8 matches of the round of 16.

Figure 1: Probability Chart of possible scores for the round of 16 of the 2024 European Championship.



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

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

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

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

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

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

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

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

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

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

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



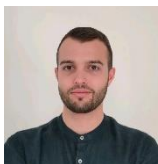
Leonardo Egidi is Assistant Professor of Statistics at the University of Trieste in Italy and a member of the AUEB Sports Analytics Group of the Athens University of Economics and Business. He has a PhD in football modeling and analytics and intense research activity in Bayesian Statistical methodology.



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 tou group <https://aueb-analytics.wixsite.com/sports>

Website tou 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](#).