

Kendall correlations and radar charts to include goals for and goals against in soccer rankings

Roy Cerqueti^{1,2} Raffaele Mattera¹ **Valerio Ficcadenti³**

¹Department of Social and Economic Sciences, Sapienza University of Rome,
Italy

²GRANEM, University of Angers, France

³School of Business, London South Bank University, United Kingdom

July 17, 2024

Introduction

Record of mathematical victory is **5 rounds in advance**: Torino (1946-1947), Fiorentina (1955-1956), Inter (2006-2007 and 2023-2024), Juventus (2018-2019), and Napoli (2022-2023).

Round	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38
2023-24 Inter Milan	W	W	W	W	W	L	W	D	W	W	W	W	D	W	W	W	W	D	W	W	W	W	W	W	W	W	W	D	W	W	D	W	W	L	W	D	D	
2022-23 SSC Napoli	W	W	D	D	W	W	W	W	W	W	W	W	W	W	L	W	W	W	W	W	W	W	W	L	W	W	L	W	D	W	D	D	W	L	W	D	W	
2018-19 Juventus FC	W	W	W	W	W	W	W	D	W	W	W	W	W	W	W	W	D	W	W	W	D	W	W	W	W	W	L	W	W	W	L	W	D	D	L	D	L	



Some literature

- ▶ Ausloos (2024) presents **new ranking indicators** for cyclists based on rank-size laws.
- ▶ Ficcadenti et al. (2023) and Ausloos (2014) view the **football rankings as unified frameworks** through a rank-size analysis.
- ▶ Sziklai et al. (2022), for an **overview of tournaments' efficacy** using Kendall correlation.
- ▶ Cerqueti et al. (2022) contains an application on football data to **rank teams according to their goals**.
- ▶ Ausloos et al. (2014) deals with the analysis of the **structure of the rankings** when considering UEFA and FIFA championships.

Our Objectives and Aims

We hypothesise a **novel scoring rule** that includes scored and conceded goals to determine the final ranking after competitions. We use it to determine the sensibility of the ranking to the **addition of performance-oriented ranking features**.

Methodology

This is a **four-step procedure** for achieving the *New Rankings*; let us look at it via football seasons.

1. Obtain **unofficial rankings** GF_r , GA_r
2. **Compute the Kendall τ** for all possible pairs of rankings
- 3.
- 4.

Correlation pairs

GF_r vs GA_r

Official Rank vs GF_r

Official Rank vs GA_r

Methodology

This is a **four-step procedure** for achieving the *New Rankings*; let us look at it via football seasons.

1. Obtain **unofficial rankings** GF_r , GA_r
2. **Compute the Kendall τ** for all possible pairs of rankings
3. **Create radar charts** and normalise the areas
4. **Detect rankings** with a target Kendall τ correlation (from the radar charts) using the official ranking \rightarrow *New Rankings*

Correlation pairs

GF_r vs GA_r

Official Rank vs GF_r

Official Rank vs GA_r

Kendall τ Correlation Analysis

The association between Official team Rankings and goal metrics-based rankings GF_r , GA_r is achieved through the Kendall τ_b variant, Kendall (1945).

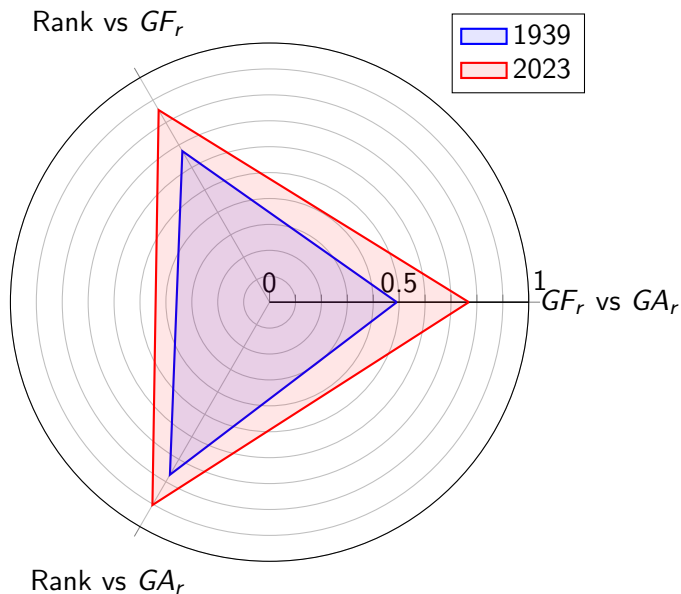
$$\tau_b = \frac{P - Q}{\sqrt{(P + Q + T)(P + Q + U)}} \quad (1)$$

Where P and Q are the number of concordant and discordant pairs, respectively, and T and U are the number of ties only in x and y , respectively.

Normalisation:

$$\tau_{b;N} = \frac{\tau_b + 1}{2}$$

Mapping Correlations into Radar Charts



Mapping Correlations into Radar Charts

Given the vertices positioned at angles θ_1 , θ_2 , and θ_3 :

$$x_h = \tau_{b;N}^{(h)} \cos(\theta_h) + \text{shift}_x$$

$$y_h = \tau_{b;N}^{(h)} \sin(\theta_h) + \text{shift}_y$$

The area (A) of the triangle is given by:

$$A = \frac{1}{2} |x_1(y_2 - y_3) + x_2(y_3 - y_1) + x_3(y_1 - y_2)| \quad (2)$$

The A s mapped back into the τ_b correlation are used as target τ_b .

Finding the *New Rankings*

To discover alternative ranking systems that reflect football championships' dynamics, including GA and GF.

1. **Generate** permutations of the positions and **calculate** $\tau_b(\textit{Official Ranking}, \textit{New Ranking})$.
2. **Identify the permutations** that match the target τ_b values.

Computational Problem

For a given number of teams, n , generate permutations to simulate various possible season outcomes. Compute Kendall τ_b for each permutation.

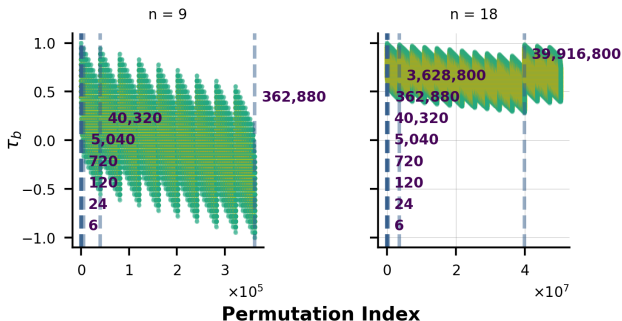


Figure: Variations in Kendall's $\tau_b^{(j)}$ correlation with permutation Index (j) for different sample sizes (n).

n	3	4	5	6	7	8	9	18	20
$n!$	6	24	120	720	5,040	40,320	362,880	6.402×10^{15}	2.432×10^{18}

Identifying Optimal Permutations

The procedure to identify optimal permutations:

1. **Extract the target** τ_b values from the geometric analysis (Areas).
2. **Calculate the absolute difference** between each τ_b target and the one resulting from j^{th} permutation.
3. **Isolate** permutations whose $\tau_b^{(j)}$ values are **nearest** to the target τ_b .

Data Summary

- ▶ Dataset includes Serie A seasons from 1930 to 2023.
- ▶ Metrics: Goals For (GF_r), Goals Against (GA_r), official rankings.

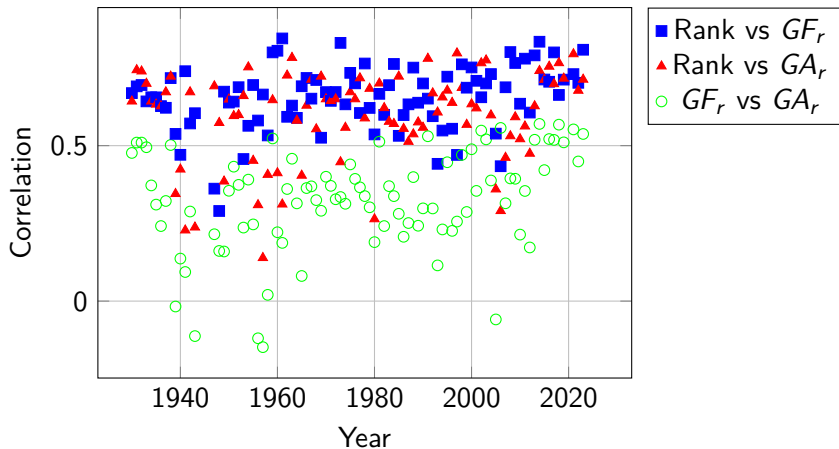


Figure: The different correlation analyses are reported over the seasons.

Results Overview - Areas time series

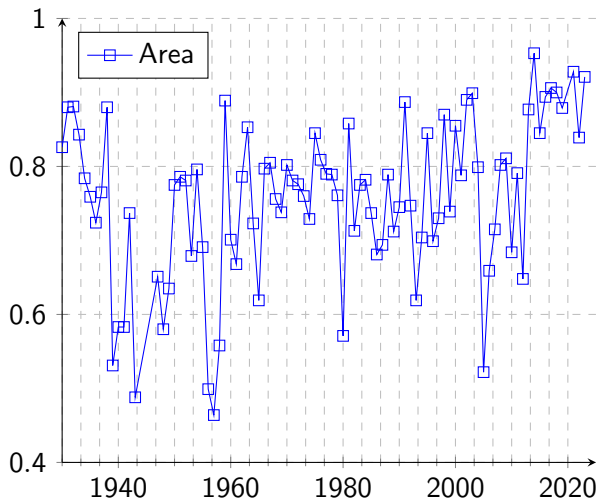


Figure: The areas calculated with Eq. (2). Each point represents a season.

Results Overview - τ_b distribution

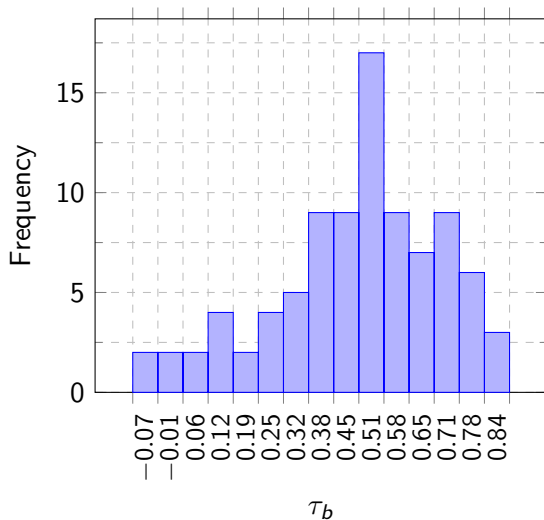


Figure: Histogram of the resulting τ_b obtained from mapping the areas back to the $[-1,1]$ correlation range - Italian Serie A history, 1930 - 2023.

Results Overview - Measure of instability?

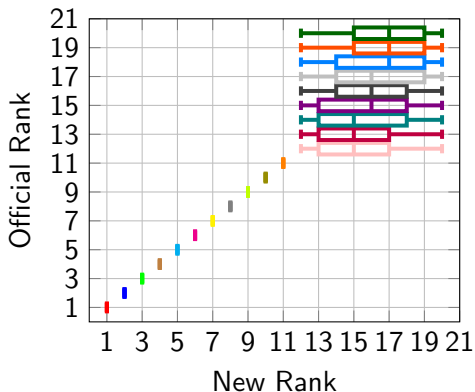


Figure: 2023 season. The target correlation is $\tau_b = 0.842962$ and the optimal correlation is $\tau_b^j = 0.842105$ with $j \in \{0, \dots, 362, 880\}$, being 362,880 the number of permutation tested in this case, when the generated permutations are stored in lexicographic order.

Results Overview - Measure of instability?

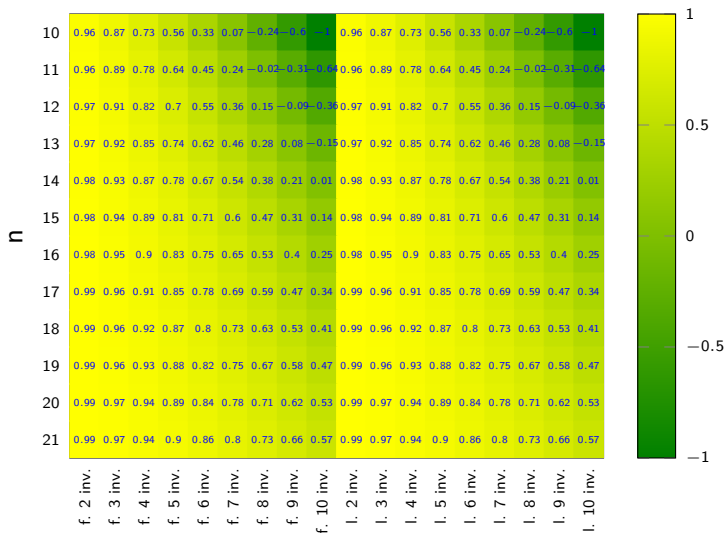


Figure: The x-axis has ticks indicating that the correlation reported in the cells is calculated comparing the original series $1, 2, \dots, n$ (y-axis) with the series where the first ('f.') or the last ('l.') k elements have been permuted, inverting their order.

Conclusion & Limitations

Conclusions:

- ▶ Presentation of a **novel approach** that integrates geometric analysis and Kendall τ
- ▶ Highlights the **multifaceted nature of competitions** and their ranking system
- ▶ **Encourages reconsideration** of conventional ranking systems thinking about the **tail of the distribution of rankings**

Limitations:





- ▶ Computational Complexity
- ▶ No single solution per each target correlation
- ▶ Generalizability, especially when more metrics are included

Future Research

- ▶ Potential to **refine ranking methodologies and integrate existing tools for comparing tournaments**, see Sum of Ranking Differences in Sziklai and Héberger (2020).
- ▶ **Extend** to other leagues and sports.
- ▶ **Incorporate additional performance metrics**, more axis on the spider chart.

Questions... ? Comments... ?
Insults... ?

References II

-  Ficcadenti, V., Cerqueti, R., & Varde'i, C. H. (2023). A rank-size approach to analyse soccer competitions and teams: the case of the Italian football league "Serie A. *Annals of Operations Research*, 325(1), 85-113.
-  Kendall, M. G. (1945). The treatment of ties in ranking problems. *Biometrika* 33(3), 239-251.
-  Sziklai, B. R., Biró, P., & Csató, L. (2022). The efficacy of tournament designs. *Computers & Operations Research*, 144, 105821.
-  Sziklai, B. R., & Héberger, K. (2020). Apportionment and districting by Sum of Ranking Differences. *PloS One*, 15(3), e0229209.