

# EUR Research Information Portal

## A new hierarchical composite indicator model for ranking the top 20 European football teams

**Published in:**  
Quality and Quantity

**Publication status and date:**  
Published: 28/02/2024

**DOI (link to publisher):**  
[10.1007/s11135-024-01842-9](https://doi.org/10.1007/s11135-024-01842-9)

**Document Version**  
Publisher's PDF, also known as Version of record

**Document License/Available under:**  
CC BY

**Citation for the published version (APA):**  
Cavicchia, C., & Sarnacchiaro, P. (2024). A new hierarchical composite indicator model for ranking the top 20 European football teams. *Quality and Quantity*, 58(5), 4033-4051. <https://doi.org/10.1007/s11135-024-01842-9>

[Link to publication on the EUR Research Information Portal](#)

### Terms and Conditions of Use

Except as permitted by the applicable copyright law, you may not reproduce or make this material available to any third party without the prior written permission from the copyright holder(s). Copyright law allows the following uses of this material without prior permission:

- you may download, save and print a copy of this material for your personal use only;
- you may share the EUR portal link to this material.

In case the material is published with an open access license (e.g. a Creative Commons (CC) license), other uses may be allowed. Please check the terms and conditions of the specific license.

### Take-down policy

If you believe that this material infringes your copyright and/or any other intellectual property rights, you may request its removal by contacting us at the following email address: [openaccess.library@eur.nl](mailto:openaccess.library@eur.nl). Please provide us with all the relevant information, including the reasons why you believe any of your rights have been infringed. In case of a legitimate complaint, we will make the material inaccessible and/or remove it from the website.



# A new hierarchical composite indicator model for ranking the top 20 European football teams

Carlo Cavicchia<sup>1</sup> · Pasquale Sarnacchiaro<sup>2</sup>

Accepted: 14 January 2024  
© The Author(s) 2024

## Abstract

The rankings of football teams are crucial to assess their sportive and non-sportive performances and allow fans and investors to have a measure of the health of the teams. The current literature lacks rankings that jointly consider the economic, sporting and popularity aspects, and our proposal aims to fill this gap. The paper proposes a new conceptual framework with the ultimate goal of building a global ranking for football teams by assessing their overall performance. Therefore, this article applies this new conceptual framework to the top 20 European teams by detecting both four specific composite indicators, corresponding to the main aspects of a football team performance, and the general composite indicator which combines the specific ones into a single index. A reflective model-based approach, based on a higher-order factor analysis, is used in order to build the composite indicators, whereas a multi-group analysis and an analysis of means are applied to test the presence of several models for the performance of football teams in the different leagues, and to test the existence of subgroups with different performances for both the scores of the general and the specific composite indicators.

**Keywords** Higher-order factor analysis · ANOM · Multi-group analysis · Football ranking · Performance

## 1 Introduction

Football is the most popular sport in the world and the wealthiest in terms of team revenue and player expenditure. By taking into account the income it generates, the number of fans it has, and the growing number of people who attend games in stadiums, it is clear that the attractiveness and the business interest aroused by this sport are constantly increasing year by year. However, although the growth in popularity is a worldwide phenomenon,

---

✉ Carlo Cavicchia  
cavicchia@ese.eur.nl

Pasquale Sarnacchiaro  
sarnacch@unina.it

<sup>1</sup> Econometric Institute, Erasmus University Rotterdam, Rotterdam, The Netherlands

<sup>2</sup> Department of Economics Management and Institution, University of Naples Federico II, Naples, Italy

the growth in business turnover mainly concerns Europe, and European football market revenue for the 2018/19 season totaled 28.9b€, which is the record of revenue generation. In detail, the revenues of the “big five European leagues” (English Premier League, Spanish La Liga, Italian Serie A, German Bundesliga and French Ligue 1) reported a growth by 10% (over 300m€) in the 2018/19 season, the second highest absolute growth considering the last ten years (Ajadi et al. 2021). However, challenging times now lie ahead for European football, and in the next seasons we may well be reporting a revenue decline. This is certainly due, at least partially, to the effects of the COVID-19 pandemic. However, in a long-term scenario, the fundamentals of the public, and hence the corporate, appetite for elite football remain strong and help the industry overcome these challenges. For example, broadcast revenues continued to be the primary driver as leagues benefited from a raise in media rights, and teams competing in European competitions profited from increased UEFA distributions.

The current situation created a different way to understand football. On the one hand, fans have the wish of winning trophies, on the other hand, shareholders and managers need to have a good economic performance. In fact, a very well managed team is that which balances the economic and sporting performances, since in the sports industry one team cannot succeed ignoring one of these two aspects (Baroncelli and Lago 2006). Therefore, it is not possible to evaluate teams’ performance taking into account only one of these two main aspects, excluding the other. This fact has been easily proven in the past, having several examples of successful teams - in terms of sporting performance—that went bankrupt. Although these two different goals can create some incompatibility, many important studies pointed out that there is a virtuous cycle between sport results and economic resources (Lago et al. 2004; Caruso 2006). In particular, economic power allowed purchasing talented players to create competitive teams and to reach good sporting results. Good sporting results directly or indirectly increase team revenues. However, this virtuous cycle can be reinforced by including another element that is distinct from the previous two but related to them: the popularity of a team. A popular team has more opportunities to explore and conquer new markets and new fans. In fact, thanks to the sale of TV rights abroad, there is an opportunity for teams to get in touch with new potential fans. A larger number of fans means more bargaining power in the distribution of revenues generated by the sale of the rights for matches and competitions. More fans also means more merchandising revenue. Determinants of the popularity of a team are the presence of great champions, victories in competitions of global interest, such as the Union of European Football Associations (UEFA) Champions League (UCL), the number of followers on social networks, and the presence in rankings compiled by various credited institutes (e.g. Deloitte). These rankings are important for different reasons: indeed, if on the one hand predicting which team will win the next Champions League is an aspect that leads to heated discussions and debates among football fans, and even attract the attention of casual watchers, on the other hand the rankings allow investors to have a measure of the health of a team. Furthermore, very often the disclosure of these rankings, accompanied by an accurate communication campaign, allows the popularity of the teams present in the rankings to increase. There are many partial rankings with respect to the object studied and territorial extension, and almost all of them exclusively consider the sports performance of the teams, whereas only a few rankings consider the economic aspects and very few ones measure the teams’ popularity. To the best of our knowledge, in the current literature there are no rankings that jointly consider the economic, sporting and popularity aspects. In this perspective, statistical tools, such as composite indicators (CoIns), represent a valid and useful instrument for the construction of rankings.

In this paper, we fill this gap and develop a conceptual framework that reflects the current strength of the football team, with the ultimate goal of building a global ranking. To this end, we consider at the same time economic performance, sports results, and the popularity of the team. In particular, we apply this new conceptual framework to the top 20 European teams present in the Deloitte Football Money League 2021 report (Ajadi et al. 2021). The resulting ranking represents an interesting addition to the well-established rankings and can be considered as a first attempt to build a global ranking of the overall performance of a football team.

The conceptual framework developed demands a specific statistical analysis strategy. In detail, the presence of different domains (i.e., first-order dimensions) that define sports and economic performance requires a second-order hierarchy to define the general composite indicator. Moreover, the method for the construction of the composite indicator needs another important property, that is, each variable belongs to a single domain. For these reasons, a second-order factor analysis is applied for the construction of the composite indicator. Furthermore, since teams are divided into leagues, it is crucial to test whether a league effect is present in the composite indicators built following the new conceptual framework. In this regard, this paper proposes to jointly consider a multi-group analysis (Cavicchia and Sarnacchiaro 2022). However, the latter analysis allows only testing of the difference between couples of groups. To overcome this problem, we propose to consider the analysis of means (Ott 1967) which, to our knowledge, has never been used in this context.

The present article is organized as follows. Section 2 introduces the conceptual framework. In Sects. 3 and 4, we briefly describe different methodologies for composite indicators' construction, and we present the statistical methods that we use to rank the teams. After the introduction of the data in Sect. 5, the case of the top 20 European teams allows us to use our methodologies in Sect. 6. Finally, we conclude the article with final comments and an outlook on future research in Sect. 7.

## 2 Conceptual framework

Many rankings are produced by several institutes and websites with different purposes. These rankings are very useful and effective for comparing teams from different countries or belonging to different leagues. However, they can only be used to compare teams in a specific aspect, such as sports performance or economic performance. This represents their main limitation because they cannot provide any evaluation of the overall performance or potential of a team. The performance of football teams is a multidimensional phenomenon that should be measured by taking into account all aspects. In the next sections, we will present a review of the most important rankings of teams provided by official and unofficial institutes and websites, and we will present our proposal of framework which includes sporting and non-sporting aspects.

### 2.1 Ranking for sports performance

The men's International Federation of Association Football (FIFA) World Ranking is a ranking system for men's national teams introduced in December 1992. Since 2018, a new version of the ranking system has been used. The latter is based on adding (or subtracting) points if teams win (or lose) a game to (or from) the previous point totals. Points which are added or subtracted are partially determined by the relative strength of the two opponents.

**Table 1** List of ranking for sports performance with related website

Name	Website
Football world ranking	<a href="http://www.clubworldranking.com">www.clubworldranking.com</a>
Fivethirtyeight	<a href="http://www.fivethirtyeight.com">www.fivethirtyeight.com</a>
Football database	<a href="http://www.footballdatabase.com">www.footballdatabase.com</a>
Soccerverse	<a href="http://www.soccerverse.com">www.soccerverse.com</a>
UEFA	<a href="http://www.uefa.com">www.uefa.com</a>
Euro Club Index	<a href="http://www.euroclubindex.com">www.euroclubindex.com</a>

Officially, there is no team-level football ranking that ranks teams worldwide, as there is for national teams. There are many unofficial rankings made by various institutes and for different purposes. The Football World Ranking project provided one of the fairest football team rankings in the world, however the project has been now discontinued. This was independent of all football and commercial organizations and was set up equally for the ATP World Tennis Ranking. Furthermore, Fivethirtyeight publishes its global football ranking for 636 international teams, which is based on a substantially revised version of ESPN's Soccer Power Index (SPI). Another interesting ranking is provided by the Football Database website where the scores of all the teams included in the survey are updated daily. Finally, the Soccerverse produces a project dedicated to the analysis of team-level football around the world. The global team ranking is the core of this project and is based on the official FIFA ranking for national teams adapted for team-level football.

Considering only European teams, an official team ranking is the one produced by UEFA which is computed on the team coefficient based on the results of the teams competing in the five previous seasons of the UCL and UEFA Europa League (UEL). The rankings determine the seeding of each team in relevant UEFA competition draws. Another extremely popular ranking is the one made by the Euro Club Index (ECI) that is a ranking of the football teams in the highest division of all European countries that shows their relative playing strength at a given point in time, and the development of playing strength in time.

Table 1 contains the complete list of ranking for sport performance with and the hypertextual reference for their websites.

## 2.2 Ranking for non-sports performance

The website Bleacherreport attempts to establish and rank the top 50 most influential teams in modern society (available at [www.bleacherreport.com](http://www.bleacherreport.com)).

Moving the focus to the rankings relating to economic aspects, an important ranking is the one provided by the website Transfermarkt. In this ranking, there are the 100 most valuable teams in the world (available at [www.transfermarkt.com](http://www.transfermarkt.com)). Another popular and accurate ranking is the one produced by the Deloitte Football Money League which is mainly focused on the economic aspects. This profiles the highest revenue-generating teams in world football and is one of the most reliable independent analyzes of the relative economic performance of the teams (Ajadi et al. 2021).

### 2.3 A new global ranking including sporting and non-sporting performance

The purpose of a ranking is to allow the reader to have concise and comparable information on the performance of the football team that can be read and understood immediately. As illustrated in the previous sections, the unofficial rankings of football teams are many and they differ for construction and objectives. The growing interest in the rankings comes from the massive increase in supply and demand for football products on a global scale. Thanks to the widespread distribution through digital channels of many football products, more and more people can choose which team or competition to follow. The strong internationalization of the sector has also broadened the horizons of possible choices. In this context, rankings on sports performance have therefore become one of the most used elements for choosing the TV product to buy as customer. However, the question naturally arises: How are the rankings of football teams constructed? The rankings can be distinguished by architecture, by method of construction of the rating and by promoter. For instance, as for the architectures, they can typically vary for:

- object studied, which can be sports performance, economic performance, popularity, or all these aspects at the same time;
- territorial extension.

Less straightforward is the understanding of the methods used for constructing the rankings. Indeed, even when only the sports performance is considered, the rankings must be built on the basis of parameters or models subjectively chosen by the provider of the rankings. These choices are needed to make the teams' performance comparable since the teams do not have opportunities for direct competition. Furthermore, the rating is rarely given by a single parameter, but it more often derives from an aggregation of many parameters. Different choices in the rating construction method (that is, selection and aggregation of parameters) could lead to a drastically different positioning of a football team in the ranking. A third criterion for distinguishing the rankings concerns the promoter which might be official or unofficial, and, it might pay more attention to a specific aspect than to others.

The aim of this paper is to propose a global ranking that includes sporting and non-sporting performance of a football team. In particular, sports performance is divided into sports results (sportive performance, SPO) achieved and quality of the game (GAM) expressed. Non-sporting performance is grouped into economic performance (ECO) and popularity (Fans, FAN). In our framework, we assume that the 4 first-order dimensions reflect the general concept to measure (i.e., sport performance). This means that if a team performs well (i.e., high sport performance), then it will also perform well in the single dimensions. In other words, we assume that the 4 dimensions are detected by disjoint groups of variables, and the correlation among the dimensions is still strong enough. Therefore, we follow a data-driven approach and want to reconstruct the relationships among the variables in order to get the maximum information from the data.

This paper can be interpreted as a first step towards the definition of a new way to understand football, such that it includes different aspects and not only considering one of them at the time. Although there are no other experiments in this direction, and thus there is no ground truth to refer to, the method used guarantees to obtain a composite indicator which complies important properties, and is suitable for the framework we propose. In detail, it is worth underlining that the method chosen for the construction of

the composite indicator and the selection of the manifest variables are perfectly aligned. The general performance of a team is reflected by 4 main dimensions which are in turn reflected by their own manifest variables. Figure 1 shows the level of correlation among the 4 dimensions and the reflective approach appears to be well motivated. All considered, the method chosen results applicable and suited for the goal of our analysis since it follows a reflective approach similarly to the hierarchical confirmatory factor analysis (HCFA, Joreskog 1969, 1978, 1979).

### 3 Aggregative composite indicators for the construction of rankings

CoIns are non-observable latent variables which are able to summarize a big amount of information according to an underlying aggregative scheme. This makes CoIns highly useful and suitable for measuring multidimensional phenomena, where it is crucial to combine different aspects in one single index, thereby preserving as much information as possible. Several methodologies, characterized by different approaches, have been proposed for the construction of CoIns throughout the years.

Factor analysis (FA, Anderson and Rubin 1956; Horst 1965) and principal component analysis (PCA, Pearson 1901; Hotelling 1933) are considered reliable methods for exploring the data and building latent factors or components. In detail, the loadings are determined taking into account the statistical relationships between the variables. For this reason, both are widely considered as weighting methods for the construction of CoIns (OECD 2004; OECD-JRC 2008). Finally, structural equation models (SEM) (Joreskog 1970; Bollen 1989; Kaplan 2000) are used in order to build a flexible system of CoIns, able to model causal relations among them. Gan et al. (2017) reported two main drawbacks with respect to the use of FA weights. The first is about the difficulty in defining the real meaning of the dimensions extracted; the second is related to the number of variables and the level of correlation among them. However, our framework guarantees to obtain 4 meaningful dimensions representing 4 theoretical domains which reflect the same general concept. Furthermore, the variables are highly correlated within each domain (SPO, GAM, ECO and FAN). Thus, even if our approach follows the HCFA, it overcomes these two potential issues.

Although CoIns are extremely valid statistical tools, they are often criticized because the methods for their construction are not always statistical and mathematically rigorous and are often based on theories that do not seem to have a solid foundation (Mazziotta and Pareto 2013). For instance when built with subjective (i.e., normative) weights and/or by following a theory which may not be respected by the data, CoIns might be therefore misleading when considered for decision-making (Nardo et al. 2005). FA instead detects the weights that best reconstruct the relationships between variables according to the estimation method chosen, therefore limiting the subjective choice of them. This approach is therefore useful when one does not have a priori knowledge about the contribution of each variable to the CoIn or when one wants to test this knowledge with the data. Nevertheless, the researcher must still choose, for instance, the number of factors to extract. Moreover, it is important to underline that when the studied latent concept presents a hierarchical structure, FA is an inappropriate method because it cannot model the hierarchical relationships among the variables, which thereby requires a model of hierarchical form (Cavicchia and Vichi 2021). A valid alternative method to construct a CoIn into the FA framework is therefore the HCFA that, however, needs to

specify beforehand the substantial information about the relations in the hierarchy. In the next Section we will introduce a methodology able build a CoIn based on a two-order hierarchy.

## 4 Methods

In this section, we present the statistical method we use to build the CoIns and the corresponding rankings for both the specific aspects and the general football team’s performance. Moreover, we briefly evoke two other methods for the data analysis. In detail, Second-Order Disjoint Factor Analysis (2O-DFA, Cavicchia and Vichi 2022) is used to build the rankings, 2O-DFA Multi-group analysis (2O-DFA-MGA, Cavicchia and Sarnacchiaro 2022) allows us to test the presence of several models for the performance of football teams in the different leagues, and finally, the analysis of means (ANOM) verifies the existence of subgroups with different performances for both the scores of the specific CoIns (SCoIns) and the score of the general CoIn (GCoIn).

### 4.1 Second-order disjoint factor analysis

2O-DFA is a factor model which consists of two orders of latent unknown constructs:  $m$  specific first-order factors and a single (nested) general factor. The model is characterized by the inclusion of some a priori substantial knowledge in the form of restrictions on the first-order loading matrix, which improves the description of the latent factors and leads to a parsimonious model with a simple loading matrix structure. Therefore, the model for centered data is defined as follows

$$\mathbf{x} = \mathbf{A}\mathbf{y} + \mathbf{w}, \tag{1}$$

where  $\mathbf{y} = \mathbf{c}g + \mathbf{u}$ . In detail,  $\mathbf{A}$  is the  $p \times m$  matrix of unknown specific first-order factor loadings,  $\mathbf{w}$  is a  $p \times 1$  random vector of errors,  $\mathbf{y}$  is the non-observable ( $m \times 1$ ) random vector denoting the first-order factor scores (i.e., SCoIns),  $g$  is a non-observable random variable normally distributed with mean 0 and variance  $\text{cov}(g) = 1$  and denoting the general factor score (i.e., GCoIn),  $\mathbf{c}$  is the  $m \times 1$  vector of unknown general factor loadings and  $\mathbf{u}$  is a  $m \times 1$  random vector of errors. It is crucial to recall that the loading matrix  $\mathbf{A}$  is restricted to the product  $\mathbf{A} = \mathbf{B}\mathbf{V}$ , where  $\mathbf{B}$  is a  $p$ -dimensional diagonal matrix and  $\mathbf{V}$  a  $p \times m$  membership matrix (i.e., row-stochastic and binary matrix). Furthermore,  $\mathbf{w} \sim N_p(\mathbf{0}_p, \Psi_{\mathbf{x}})$ , where  $\text{cov}(\mathbf{w}) = \Psi_{\mathbf{x}}$  is the  $p$ -dimensional diagonal positive definite variance-covariance matrix of the errors of the first-order model,  $\mathbf{u} \sim N_m(\mathbf{0}_m, \Psi_{\mathbf{y}})$ , and where  $\text{cov}(\mathbf{u}) = \Psi_{\mathbf{y}}$  is the  $p$ -dimensional diagonal positive definite variance-covariance matrix of the errors of the second-order model. Furthermore, it is assumed that  $\mathbf{y} \sim N_m(\mathbf{0}_m, \Sigma_{\mathbf{y}})$ , where  $\Sigma_{\mathbf{y}}$  is the correlation matrix of the first-order factors, that the errors in the two models are uncorrelated  $\text{cov}(\mathbf{w}, \mathbf{u}) = \mathbf{0}$ , and that the errors and factors are uncorrelated, that is,  $\text{cov}(\mathbf{u}, g) = \mathbf{0}$ .

Given these assumptions, it can be derived that  $\mathbf{x} \sim N_p(\mathbf{0}_p, \Sigma_{\mathbf{x}})$ , where the variance-covariance matrix  $\Sigma_{\mathbf{x}}$  is

$$\Sigma_{\mathbf{x}} = \mathbf{B}\mathbf{V}(\mathbf{c}\mathbf{c}' + \Psi_{\mathbf{y}})\mathbf{V}'\mathbf{B} + \Psi_{\mathbf{x}}, \tag{2}$$

where

$$\Sigma_{\mathbf{y}} = \mathbf{c}\mathbf{c}' + \Psi_{\mathbf{y}}, \tag{3}$$



such that

$$\mathbf{V} = (\mathbf{v}_{jq}) \text{ where } \mathbf{v}_{jq} \in \{0, 1\} \quad (4)$$

$$\mathbf{V}\mathbf{1}_m = \mathbf{1}_p \quad (5)$$

$$\mathbf{B} = \text{diag}(b_1, \dots, b_p) \text{ with } b_j^2 > 0 \quad (6)$$

$$\mathbf{V}'\mathbf{B}\mathbf{B}\mathbf{V} = \text{diag}(b_{\cdot 1}^2, \dots, b_{\cdot m}^2) \text{ with } b_{\cdot q}^2 = \sum_{j \in C_q} b_j^2, \quad (7)$$

where  $C_q$ , with  $q = 1, \dots, m$  representing the  $q$ th group of variables, and  $\mathbf{1}_m$  and  $\mathbf{1}_p$  denote the unit vectors of dimensions  $m$  and  $p$ , respectively. Let us consider a random sample of  $n > J$  multivariate observations  $\mathbf{x}_i = [x_{i1}, \dots, x_{ij}]'$ ,  $i = 1, \dots, n$  drawn of  $\mathbf{x}$ , with mean vector  $\bar{\mathbf{x}}$ , and  $p$ -dimensional variance-covariance matrix  $\mathbf{S} = \frac{1}{n} \sum_{i=1}^n \mathbf{x}_i \mathbf{x}_i'$ , the model in matrix form then corresponds to

$$\mathbf{X} = \mathbf{g}\mathbf{c}'\mathbf{B}\mathbf{V}' + \mathbf{E}, \quad (8)$$

where  $\mathbf{X} = [\mathbf{x}_1, \dots, \mathbf{x}_n]'$  is the  $n \times p$  matrix containing the  $n$  multivariate observations,  $\mathbf{g} = [g_1, \dots, g_n]'$  is the non-observable  $n \times 1$  vector denoting the second-order (general) factor scores, and  $\mathbf{E} = \mathbf{U}\mathbf{B}\mathbf{V}' + \mathbf{W}$  is the  $n \times p$  matrix of errors. In detail,  $\mathbf{W} = [\mathbf{w}_1, \dots, \mathbf{w}_n]'$  with dimensions  $n \times p$  and  $\mathbf{U} = [\mathbf{u}_1, \dots, \mathbf{u}_n]'$  with dimensions  $n \times m$  are matrices containing the non-observable errors related to first-order and second-order models, respectively.

The maximization of the reduced log-likelihood (i.e., conditional on  $\boldsymbol{\mu}$  equal to the sample mean) corresponds to the minimization of the discrepancy function

$$D(\mathbf{B}, \mathbf{V}, \mathbf{c}, \boldsymbol{\Psi}_x, \boldsymbol{\Psi}_y) = \log |\boldsymbol{\Sigma}_x| + \text{tr}(\boldsymbol{\Sigma}_x^{-1}\mathbf{S}). \quad (9)$$

2O-DFA reconstructs  $\boldsymbol{\Sigma}_x$  in terms of  $2p + m$  unknown free parameters in  $\mathbf{B}$ ,  $\mathbf{V}$ ,  $\boldsymbol{\Psi}_x$  and  $\mathbf{c}$ , which are estimated according to a cyclic block coordinate descent algorithm.

## 4.2 2O-DFA multi-group analysis

Cavicchia and Sarnacchiaro (2022) propose 2O-DFA-MGA that tests the potential differences between groups in the definition of the general construct in 2O-DFA. The procedure extends the one proposed by Henseler et al. (2009) called the PLS-MGA procedure, which is also implemented in SmartPLS.

In detail, 2O-DFA-MGA tests potential differences on 2O-DFA second-order coefficients (i.e., the loading between SCoIns and GCoIn) and allows verifying if the groups taken into account define the general construct according to the same chosen model (i.e., the SCoIns considered) equal to the one considered to measure it in the entire sample. Multi-group analyses usually imply the testing of measurement model invariance (Millisap 2011; Henseler et al. 2016), but Cavicchia and Sarnacchiaro (2022) do not consider it necessary since the assumption that the membership matrix  $\mathbf{V}$  is equal across groups guarantees compliance to equal-form measurement model invariance. 2O-DFA-MGA implements a bootstrap procedure which requires the following: (i) to select the coefficient to test (i.e.,  $c$ : an element of vector  $\mathbf{c}$ ), (ii) to run bootstrapping for each group considered

in the multi-group analysis, (iii) to compare and evaluate the observed distribution of the bootstrap results - similarly to a Mann–Whitney–Wilcoxon test (Mann and Whitney 1996; Wilcoxon 1947) - according to the conditional probability  $P(c_s^1 > c_s^2 | c^1 \leq c^2)$ , where  $c_s^1$  and  $c_s^2$  represent coefficient estimates, while  $c^1$  and  $c^2$  represent the true population parameters of groups 1 and 2, respectively, (iv) if the probability is bigger than 0.90 or smaller than 0.10, it means that there is a significant difference of the group-specific coefficients. Cavicchia and Sarnacchiaro (2022) suggest considering a minimum number of bootstrapping permutation runs equal to 10,000.

The main limitation of 2O-DFA-MGA is that it only allows for the comparison of two groups. This means that, when the groups are more than two, this procedure only allows a one-versus-all approach.

### 4.3 Analysis of means

ANOM, introduced by Ott (1967), is a graphical method to compare means, rates and proportions among treatment groups to see if any of them differs significantly from the grand mean. This method can be used with both balanced and unbalanced data (Pallmann and Horthon 2016). One of the outcomes of ANOM is the so-called ANOM chart which is analogous to a control chart portrays decision lines in such a way that both statistical significance and the practical significance of samples may be assessed simultaneously.

ANOM is used as an alternative to the analysis of variance (ANOVA) or in a complementary way. However, it is worth observing that ANOM provides additional information with respect to the ANOVA test, particularly when the null hypothesis is rejected. In fact, ANOM not only reveals statistically significant differences among treatment groups, but also identifies those populations that lead to these differences. Moreover, ANOM can be considered a special case of a much broader statistical concept known as multiple contrast tests (Bretz et al. 2001).

Another interesting aspect to observe is that observational studies are often unbalanced, and in this case, the ANOM test procedure is slightly more complicated. The complication is due to the fact that the decision lines around the grand mean will be tighter for the larger samples and wider for the smaller samples. As a result, the decision lines for studies with unequal samples are expressed as follows (Nelson 1983)

$$UDL, LDL = \bar{Y} \pm m_\alpha \sqrt{MS_e \frac{N - n_i}{N \times n_i}}, \quad (10)$$

where  $\bar{Y}$  is the grand mean,  $m_\alpha$  is the tabulated quantile of an unbalanced design for a particular significance level  $\alpha$  (for tables, see Nelson et al. 2003, 2005),  $n_i$  is the number of individuals in the  $i$ th groups,  $N$  is the total number of individuals in all groups and  $MS_e$  is a pooled variance.

## 5 Data

Football teams' performance is a multidimensional phenomenon which can be described by a huge quantity of information and statistics. To measure it, it is crucial to find a way to take into account all the specific aspects that constitute the phenomenon. Statistics related to economic performance, the number of fans, sports performance, and the quality of the

**Table 2** List of variables

Aspect	Label	Name	Aspect	Label	Name
ECO	Rev	Revenue	SPO	Leag	National league performance
ECO	CRev	Commercial Revenue	SPO	Per	All competitions performance
ECO	BRev	Broadcast Revenue	SPO	Ppg	Points per game
ECO	MRev	Matchday Revenue	GAM	Gpg	Goals per game
FAN	Fac	Facebook followers	GAM	Poss	Possession (%)
FAN	Inst	Instagram followers	GAM	Pass	Pass success (%)
FAN	Twit	Twitter followers	GAM	Spg	Shots per game
FAN	Yout	Youtube followers	GAM	Sot	Shots on target per game
FAN	TikT	TikTok followers	GAM	Dpg	Dribbling per game

**Table 3** List of teams

Name	Abbreviation	Name	Abbreviation
Barcelona	BAR	Arsenal	ARS
Real Madrid	RMA	Borussia Dortmund	BDO
Bayern Munchen	BMO	Atletico Madrid	AMA
Manchester United	MUN	Inter	INT
Liverpool	LIV	Zenit	ZEN
Manchester City	MCI	Schalke 04	SCH
Paris Saint German	PSG	Everton	EVE
Chelsea	CHE	Lyon	LYO
Tothenam	TOT	Naples	NAP
Juventus	JUV	Eintracht Frankfurt	EFR

game are expanding every year, and the need to build an aggregated index to monitor the teams' performance is even more important, for both fans and investors.

After the preliminary steps of the analysis, the final dataset considered in this paper consists of 18 variables (Table 2) and 20 observations (i.e., the top 20 European football teams with the highest revenue for the season 2019/20, Table 3). The choice to consider the top 20 European football teams is motivated by the need to have a clear and fair comparison with the ranking provided by Deloitte and to show that the inclusion of other dimensions (SPO, GAM, and FAN) plays a crucial role in the definition of the general performance indicator. The variables in the dataset come from the Deloitte Football Money League 2021 report (Ajadi et al. 2021), the official website of each league, and the website [www.whoscored.com](http://www.whoscored.com). The variables are regularly updated and available for free on their websites. It is worth observing that, following the conceptual framework presented in Sect. 2, the variables belong to the aspects as reported in Table 2.

In order to measure SPO, we consider two variables ("National league performance" and "All competitions performance") that cannot be directly observed. It is worth explaining how we compute these two variables. First, it is important to note that there are several competitions that have different impact in the season of the respective team and that the priority of the team can also differ with respect to the strategic planning of the season. Furthermore, the national leagues are different (e.g., the number of the teams in the league

might differ from one league to another), and it is thereby necessary to have an index which also takes into consideration these features. “National league performance” is computed following the approach proposed by Szymanski and Kuypers (1999):

$$\text{Leag} = -\log\left(\frac{t}{k-t+1}\right), \quad (11)$$

where  $k$  is the number of teams who take part in the competition and  $t$  is the position they achieved in the ranking. Whereas “All competitions performance” is obtained following the methods used by Barajas et al. (2005) who defined an aggregated index with the aim of differentiating the competitions and their impact:

$$\text{Per} = \text{Cup Pts} + 2 \times \text{UEL Pts} + 2 \times \text{League Pts} + 3 \times \text{UCL Pts}, \quad (12)$$

where Cup Pts represents the points obtained by a team in the national cup, UEL Pts and UCL Pts represent the points for the UEFA Europa League (UEL) and the UCL, respectively, and finally League Pts represents the points obtained in the national league. In detail, Cup Pts is computed according to the measurement system used by Barajas et al. (2005), who developed a diagram where each stage represents a different level of points for the teams (39 points for the winner, 33 points for the runner-up, and so on). UEL Pts and UCL Pts are computed according to the following rules: the group stage will be awarded with the number of points according to the UEFA criteria; that is, 3 points for a victory, 1 for a draw and 0 for defaults, then the round-of-16 guarantees 6 points, the quarter-finals 7, the semi-finals 8 points, and the final 9 points. Eventually, the system also allocates 3 points for each victory and 1 for a draw during the knock-out phases.

## 6 Results

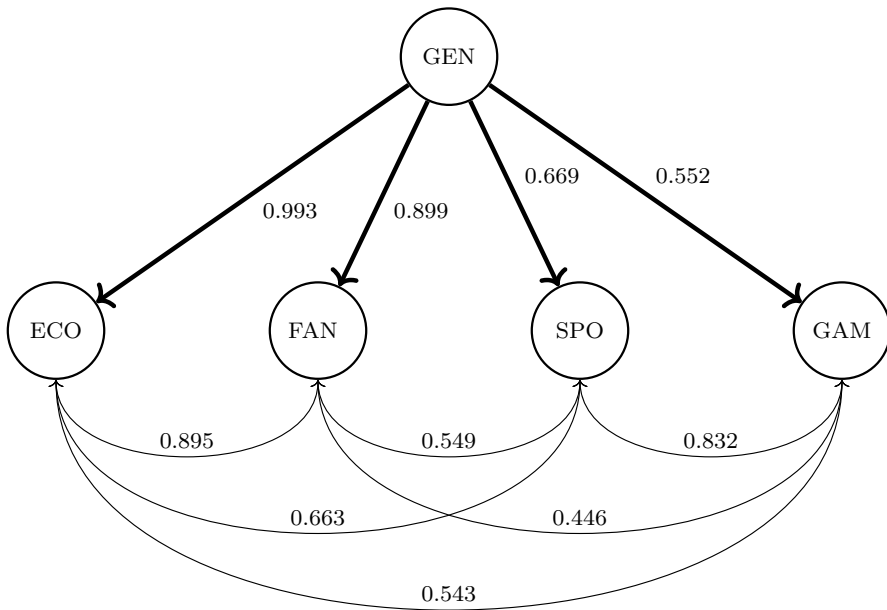
In this section we present three main analyses: the general model defining the GCoIn for football teams’ performance, the multigroup analysis and the analysis of means. In Sect. 6.1, 2O-DFA is applied on the dataset described in Sect. 5 after standardizing the variables to  $z$ -scores,<sup>1</sup> whereas in Sect. 6.2 we carry out 2O-DFA-MGA in order to test the national effect in the definition of our general CoIn. In detail, we test the model for the British teams versus all the others. Finally, in Sect. 6.3 we investigate the possible presence of statistically different subgroups for the global and specific rankings.

### 6.1 The proposed ranking via a model-based CoIn

In order to model the conceptual framework, we present a second-order hierarchy that best represents the football teams’ performance as a whole.

Aware that a possible misspecification of the measurement model can create measurement error, which in turn affects the estimation of SCoIns and CoIns (Jarvis et al. 2003), in our research particular attention is given to the construction and validation of the measurement models (i.e. the model that specifies the relationships between SCoIns) and the

<sup>1</sup> 2O-DFA is scale invariant, thus the model results equivariant under an affine transformation of the data. In this application, variables are standardized, yet the re-scaling normalization could also be used obtaining exactly the same results (Cavicchia and Vichi 2022)



**Fig. 1** Path diagram of the proposed GCoIn with second-order loadings and correlations among SCoIns

corresponding manifest variables. Depending on the causal priority between the manifest variables and the latent variable (Bollen 1989), the first choice to take for the specification of the measurement model is formative or reflective. Starting from the four main theoretical decision rules proposed by Jarvis et al. (2003) and taking into account the guidelines and suggestions introduced by Diamantopoulos and Sigauw (2006) and Maggino (2017), we opt for a reflective model specification. This choice is based on considerations that latent constructs (i.e., GCoIn and SCoIns) are existing independently from the measures used (i.e., variables), and the direction of causality is from latent constructs to variables (theoretical considerations). Next, to corroborate the suitability of the chosen model, some empirical tests are carried out: variable intercorrelation and variable relationships with latent construct antecedents and consequences (empirical considerations).

In particular, we build 4 SCoIns: ECO, FAN, SPO and GAM, for the definition of the main aspects of football teams' performance. All the factor loadings of the model result statistically significant, while all the variables are normally distributed ensuring that the distributional assumptions of 2O-DFA are met. The four SCoIns are reliable, and all SCoIns but GAM are unidimensional. In detail, the reliability is assessed through Cronbach's  $\alpha$  (Cronbach 1951)—all the SCoIns have  $\alpha$  larger than 0.90—and the unidimensionality is measured through the second largest eigenvalue of the variance-covariance sub-matrices of variables related to the SCoIn (Cavicchia and Vichi 2021)—three out of four have unidimensionality lower than 0.4, whereas GAM has unidimensionality equal to 1.02, which is anyway very close to 1. The unidimensionality and reliability of measurement models of the four SCoIns justify the choice of a reflective model for them.

Figure 1 displays the hierarchy, and it is possible to see that ECO is the most important SCoIn in the definition of the GCoIn (GEN), with loading equal to 0.993, whereas FAN, SPO and GAM have less importance with loadings equal to 0.899, 0.669 and 0.552, respectively. The model therefore shows how the economic performance (ECO) is crucial

**Table 4** Pearson's correlations between the 4 SCoIns and the GCoIn

	ECO	FAN	SPO	GAM
GEN	0.941	0.905	0.818	0.745

**Table 5** Rankings based on the 4 SCoIns and on the GCoIn

ECO	FAN	SPO	GAM	GEN
BAR	BAR	BMO	MCI	BAR
RMA	RMA	PSG	BMO	RMA
BMO	MUN	LIV	PSG	BMO
MUN	LIV	JUV	BAR	PSG
LIV	PSG	RMA	JUV	LIV
PSG	JUV	INT	BDO	MCI
MCI	BMO	MCI	LIV	MUN
TOT	CHE	BAR	NAP	JUV
CHE	MCI	ZEN	RMA	CHE
ARS	ARS	BDO	CHE	BDO
JUV	TOT	MUN	INT	INT
BDO	AMA	AMA	ZEN	TOT
AMA	BDO	CHE	MUN	ARS
INT	INT	NAP	LYO	AMA
SCH	EVE	TOT	EFR	ZEN
EVE	NAP	LYO	TOT	NAP
ZEN	LYO	ARS	ARS	LYO
LYO	ZEN	EFR	EVE	EFR
EFR	SCH	EVE	SCH	EVE
NAP	EFR	SCH	AMA	SCH

in the definition of the overall performance of a football team, although it should be underscored that FAN has a role that is anything but marginal. The correlations among the four SCoIns justify the choice of a reflective model for GEN; their values show that they share a common information which is summarized by GEN. In particular, ECO and FAN have a correlation equal to 0.895 that confirms that a high number of fans leads to higher revenue, while the correlation between SPO and GAM shows that good sportive performance arrives through a good quality of game expressed.

Table 4 shows the correlations between GEN and the four main SCoIns of the performance of the football teams, confirming the consistency of our conceptual framework. These correlations follow the same order of the loadings displayed in Fig. 1, and underline that ECO is the most important SCoIn in the definition of GEN and their final rankings result being the most similar (Table 5). The differences are explained by the impact that FAN, SPO, and GAM have on GEN. For instance, BAR is the team with the highest revenue and the one with the highest number of fans around the world; it therefore resulted as first in the ranking of ECO and FAN. These two important results make BAR first also in the general ranking. The same conclusion might be done for RMA as second in the rankings of ECO, FAN and GEN. SPO is measured in terms of victories in the different national and European competitions, where BMO is therefore the best team of the

**Table 6** 2O-DFA-MGA: comparison among second-order loadings

SCoIn	British teams	Others	2O-DFA-MGA <i>p</i> -value
ECO	0.780	0.999	0.403
FAN	0.827	0.924	0.364
SPO	0.962	0.702	0.900*
GAM	0.835	0.571	0.869

\*Denotes *p*-values smaller than 0.10 or bigger than 0.90

2019/2020 season, since it won all the competitions it participated in. PSG finishes second in the SPO ranking, being the runner-up of UCL and the winner of Ligue 1 and Coupe de France (the French national cup). An interesting epitome is MCI, which follows the seventh, ninth and seventh positions in the rankings of ECO, FAN and SPO, respectively. The first two results might be explained by the fact the MCI was purchased by Abu Dhabi United Group only in 2008, with the goal of making the team one of the best in the world after a long story with few highs and many lows. MCI was for a long time the less attractive and winning team in Manchester; however, the GAM ranking shows that MCI is the best team in terms of quality of the game. The latter is given by the fact that the team is led by one of the most spectacular and winning coaches in football history, Pep Guardiola. This important result makes MCI gain the sixth position within the GEN ranking, right after LIV which performs better in ECO, FAN and SPO. Another interesting example is NAP which results last in ECO and in the second half of the rankings for FAN and SPO. Particularly, the good performance in GAM (i.e., eighth) allows NAP to climb four positions in GEN compared with the ranking in ECO.

The ranking determined by GEN is compared with the ranking provided by UEFA. Although, it is worth stressing that the variables used in this analysis refer to the season 2020/2021 only, and therefore the comparison with the ranking by UEFA remains difficult to interpret, the rank correlation between GEN and the one provided by UEFA is equal to 0.815. This shows that the introduction of new aspects (e.g., ECO, FAN and GAM) for assessing the clubs' performance is needed, however the high correlation we found shows the robustness of our framework.

## 6.2 Multigroup analysis for testing the national effect

2O-DFA-MGA is applied in order to test the national effect in the definition of GEN. The aim of this multigroup analysis consists of assessing a statistical significant difference among the second-order loadings estimates. In detail, we analyze the British teams compared to all the others, which are hereby considered as an unique subgroup.

The second-order loadings estimates is different for the model estimated for the British teams and for the model used for the other teams. The latter model results very similar to the general one, whereas the former seems to attribute more importance to SPO. In addition, ECO results to be the least important in the definition of GEN (Table 6). These results might be due to the fact that the sample size for the British teams is only 7, therefore, it is important to test the validity of these results according to the procedure proposed by Cavicchia and Sarnacchiaro (2022). After running 2O-DFA to the subgroups separately, in order to test whether the general model for the definition of GEN is different for the British

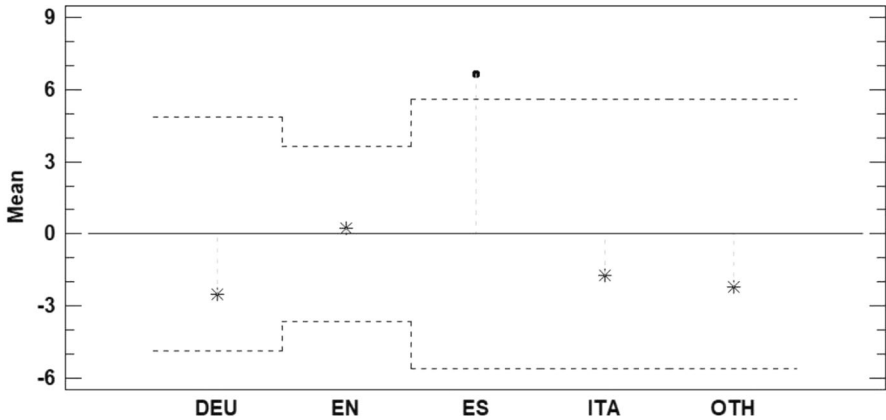


Fig. 2 ANOM chart for the scores of FAN

teams and for the others, the procedure is established to run 10,000 bootstrapping. Only one loading results being statistically different, i.e. the one regarding SPO (Table 6). The latter result is not very strong (i.e., significant at a 90% significance level and very close to the threshold), and it does not allow us to conclude that the model for British teams defines GEN differently.

2O-DFA-MGA allowed detecting an interesting result about the definition of the general model which results reliable, however a deeper investigation about the scores obtained for the four aspects is needed.

### 6.3 Analysis of means for the subgroup differences

After verifying that the global scoring system is suitable for all teams, the presence of statistically different sub-groups for the GCoIn and SCoIns is explored. The subgroups considered in this study are the ones composed by British (EN), Spanish (ES), Italian (ITA), German (DEU) and French and Russian teams (OTH) taken together. Therefore, a post-hoc analysis can initially be conducted to determine which groups have means that are significantly different and which do not. Particularly, in order to check statistically significant differences, a pairwise comparison (Hsu 1996) is made with the Tukey-Kramer test for an unequal number of samples in each group (i.e., unbalanced design). For all computations, the significance level  $1 - \alpha = 0.95$  is used. The results show a difference in the average of Spain compared to that of some other countries, especially for the variables FAN, and, subsequently, ECO (Table 7).

Finally, to investigate the existence of differences in the averages among nations for both the SCoIns and GEN, the ANOM graphs are constructed. The graph shows the mean of each factor level, the grand mean, and the decision limits. If a point falls outside the decision limits, then evidence exists that the factor level mean represented by that point is significantly different from the grand mean. All the charts built for SCoIns and GEN show that the group effects are well within the decision limits, which means that there is no evidence of interaction. The only SCoIn that highlights a group with a mean different from the grand mean is that of FAN for Spanish teams (Fig. 2).



**Table 7** Comparisons for SCoIns and GCoIn: a multiple comparison procedure. \* denotes statistically significant differences between any pair of means at the confidence level  $1 - \alpha = 0.95$ 

Contrast	ECO	FAN	SPO	GAM	GEN
DEU - EN	-2.218	-.2735	-0.327	0.413	1.437
DEU - ES	-4.703	-9.117*	-2.231	0.290	0.970
DEU - ITA	1.030	-0.771	-1.864	-1.990	1.233
DEU - OTH	0.717	-0.304	-1.857	-2.000	1.247
EN - ES	-2.485	-6.383*	-1.903	-0.123	-0.467
EN - ITA	3.248	1.964	-1.537	-2.403	-0.204
EN - OTH	2.935	2.430	-1.530	-2.413	-0.191
ES - ITA	5.733*	8.347*	0.367	-2.280	0.263
ES - OTH	5.420*	8.813*	0.373	-2.290	0.277
ITA - OTH	-0.313	0.467*	0.007	-0.010	0.013

## 7 Discussion and conclusion

The world of football has undergone profound changes in the last twenty years, recording a significant growth in spectators and in the income generated by football leagues and teams. This trend created a strong interest from consulting firms and investment funds which saw the world of football as a happy island to focus on in order to further develop their businesses. The new investors shifted their attention from the sports scores of the teams to their economic performance, pushing the football teams to maintain healthy balance sheets. An example of this is the so-called financial fair play proposed by UEFA, which European teams must respect. To intercept new business opportunities around the world, new organizational models were proposed with greater managerial and strategic skills. This new context favored internationalization strategies developed by leagues and teams, which led to a widespread and global interest in the main tournaments present in the various continents. And in this, UEFA Champions League represents one of the most important epitomes. This transformation process gave rise to a strong interest in creating rankings capable of comparing teams from different nations and continents.

Various unofficial—and in any case partial—rankings were therefore constituted by taking into account their objective and their territorial extension. In this paper, an initial response was given to the need to build partial and general rankings of European football teams. Specifically, after a careful examination of the literature, a framework was identified consisting of an overall indicator of the health of each team and four specific indicators relating to the four identified aspects: Economic Performance, Popularity, Sports Performance, and Game Quality. This conceptual framework was analyzed through a statistical model of higher-order indicators that made it possible to have both specific indicators for individual aspects relating to the performance of football teams, and an overall indicator capable of measuring the health status of a club in a given moment. This higher-order model showed that all the four aspects considered did contribute significantly to the construction of the general indicator. However, the economic aspects and popularity of a club were predominant in determining the general ranking. This indicator model was applied to the top 20 European teams with the highest revenue, and our results confirmed the validity of the proposal. Barcelona and Real Madrid were the top two teams in the general ranking and in the partial rankings related to economic performance and popularity.

Therefore, our statistical strategy results are exploitable by different subjects. On the one hand, scholars who, in their research, wonder how to synthesize different CoIns (and

their rankings) made on the same statistical observations for different competitions could be interested in our approach. In fact, this involves the construction of a CoIn that, based on a defined model, estimates the weights following the criterion of the information contained in the data. On the other hand, investors and economic operators in the football sector that, considering our global approach, may be tempted to evaluate the performance of a football team as the union of several aspects that influence each other, going beyond the evaluation of single variables. This aspect results crucial for studying multidimensional phenomena that are characterized by several aspects.

Finally, the complementary use of 2O-DFA-MGA and ANOM was proposed both to test the validity of the model for all football teams considered and to evaluate the presence of any subgroups among the football teams. In the first case, it was found that the proposed model with the respective coefficients could be valid for all teams, whereas in the second analysis it was verified that the Spanish teams had on average a greater popularity than the others. The results we reached are encouraging and push us to follow two possible lines of development: on the one hand, an application of the identified model to a larger sample with extra-European teams, and, on the other hand, a time varying study of the model.

**Funding** The authors have not disclosed any funding.

## Declarations

**Conflict of interest** The authors declare that no funds, grants, or other support were received during the preparation of this manuscript. The authors declare that they have no conflict of interest.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## References

- Ajadi, T., Bridge, T., Hanson, C., Hammond, T., Udwardia, Z.: Deloitte Football Money League 2021. Tech. rep. Manchester, UK: Deloitte Sports Business Group. (2021). <https://www2.deloitte.com/content/dam/Deloitte/uk/Documents/sports-business-group/deloitte-uk-deloittefootball-money-league-2021.pdf>
- Anderson, T., Rubin, H.: Statistical inferences in factor analysis. Proceedings of the Third Symposium on Mathematical Statistics and Probability **5**, 111–150 (1956)
- Barajas, A., Fernández-Jardón, C., Crolley, L.: Does sports performance influence revenues and economic results in Spanish football? MPRA Paper 3234. University Library of Munich, Germany (2005). <https://doi.org/10.2139/ssrn.986365>
- Baroncelli, A., Lago, U.: Italian Football. *J. Sports Econ.* **7**(1), 13–28 (2006)
- Bollen, K.: Structural Equations with Latent Variables. Wiley, New York (1989)
- Bretz, F., Genz, A., Hothorn, L.: On the numerical availability of multiple comparison procedures. *Biom. J.* **43**, 645–656 (2001)
- Caruso, R.: Umberto Lago, Alessandro Baroncelli, Stefan Szymanski (a cura di), Il business del calcio, successi sportivi e rovesci finanziari, Egea, Milano, 2004. *Rivista di Diritto ed Economia dello Sport* **2**(2), 123–129 (2006)
- Cavicchia, C., Sarnacchiaro, P.: A multi-group higher-order factor analysis for studying the gender-effect in teacher job satisfaction. *METRON* **80**, 23–38 (2022). <https://doi.org/10.1007/s40300-021-00204-x>

- Cavicchia, C., Vichi, M.: Statistical model-based composite indicators for tracking coherent policy conclusions. *Soc. Indic. Res.* **156**(2), 449–479 (2021)
- Cavicchia, C., Vichi, M.: Second-order disjoint factor analysis. *Psychometrika* **87**(1), 289–309 (2022)
- Cronbach, L.: Coefficient alpha and the internal structure of tests. *Psychometrika* **16**(3), 297–334 (1951)
- Diamantopoulos, A., Sigauw, J.: Formative Versus Reflective Indicators in Organizational Measure Development: A Comparison and Empirical Illustration. *Br. J. Manag.* **17**, 263–282 (2006)
- Gan, X., Fernandez, I., Guo, J., Wilson, M., Zhao, Y.: When to use what: Methods for weighting and aggregating sustainability. *Ecol. Ind.* **81**, 491–502 (2017)
- Henseler, J., Ringle, C., Sarstedt, M.: Testing measurement invariance of composites using partial least squares. *Int. Mark. Rev.* **33**(3), 405–431 (2016)
- Henseler, J., Ringle, C., Sinkovics, R.: The use of partial least squares path modeling in international marketing. In: *Handbook of Partial Least Squares: Concepts, Methods and Applications*. Emerald Group Publishing Limited, Bingley, pp. 277–319 (2009)
- Horst, P.: *Factor Analysis of Data Matrices*. Holt, Rinehart and Winston, New York (1965)
- Hotelling, H.: Analysis of a Complex of Statistical Variables Into Principal Components. *J. Educ. Psychol.* **24**(6), 417–441 (1933). (and 498–520)
- Hsu, J.: *Multiple Comparisons*. Chapman & Hall, London (1996)
- Jarvis, C., Mac Kenzie, S., Podsakoff, P.: A critical review of construct indicators and measurement model misspecification in marketing and consumer research. *J. Consum. Res.* **30**, 199–218 (2003)
- Joreskog, K.: A general approach to confirmatory maximum-likelihood factor analysis. *Psychometrika* **34**(2), 183–202 (1969)
- Joreskog, K.: A general method for analysis of covariance structure. *Biometrika* **57**, 239–251 (1970)
- Joreskog, K.: Structural analysis of covariance and correlation matrices. *Psychometrika* **43**(4), 443–477 (1978)
- Joreskog, K.: A general approach to confirmatory maximum likelihood factor analysis with addendum. In: Jöreskog, K.G., Sörbom, D. (eds.) *Advances in Factor Analysis and Structural Equation Models*. Cambridge (1979)
- Kaplan, D.: *Structural Equation Modeling: Foundations and Extensions*. SAGE Publications Inc, Thousand Oaks (2000)
- Lago, U., Baroncelli, A., Szymanski, S.: *Il business del calcio: successi sportivi e rovesci finanziari*. Egea, Milano (2004)
- Maggino, F.: Developing indicators and managing the complexity. In: *Complexity in Society: From Indicators Construction to their Synthesis*. Springer, Berlin, pp. 87–114 (2017)
- Mann, H., Whitney, D.: On a test of whether one of two random variables is stochastically larger than the other. *Ann. Math. Stat.* **18**(1), 50–60 (1996)
- Mazziotta, M., Pareto, A.: Methods for constructing composite indices: Pareto for all or all for one? *Rivista Italiana di Economia Demografia e Statistica* **67**(2), 67–80 (2013)
- Millsap, R.: *Statistical Approaches to Measurement Invariance*. Routledge, New York (2011)
- Nardo, M., Saisana, M., Saltelli, A., Tarantola, S.: *Tools for Composite Indicators Building*. Report EUR 21682. European Commission (Join Research Centre, Ispra, Italy) (2005)
- Nelson, P.: A comparison of sample sizes for the analysis of means and the analysis of variance. *J. Qual. Technol.* **15**, 33–39 (1983)
- Nelson, P., Coffin, M., Copeland, K.: *Introductory Statistics for Engineering Experimentation*. Academic Press, New York (2003)
- Nelson, P., Wludyka, P., Copeland, K.: *The Analysis of Means*. Society for Industrial and Applied Mathematics (2005)
- OECD: *The OECD-JRC Handbook on Practices for Developing Composite Indicators*, Paper Presented at the OECD Committee on Statistics. OECD Publisher, Berlin (2004)
- OECD-JRC: *Handbook on Constructing Composite Indicators. Methodology and User Guide*. OECD Publisher, Paris (2008)
- Ott, E.: Analysis of means: a graphical procedure. *J. Qual. Technol.* **24**, 101–109 (1967)
- Pallmann, P., Horthon, L.: Analysis of means: a generalized approach using R. *Ind. Qual. Control* **43**(8), 1541–1560 (2016)
- Pearson, K.: On lines and planes of closest fit to systems of points in space. *Phil. Mag.* **2**(11), 559–572 (1901)
- Szymanski, S., Kuypers, T.: *Winners and Losers: The Business Strategy of Football*. Penguin Group, England (1999)
- Wilcoxon, F.: Probability tables for individual comparisons by ranking methods. *Biometrics* **3**(3), 119–122 (1947)

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.