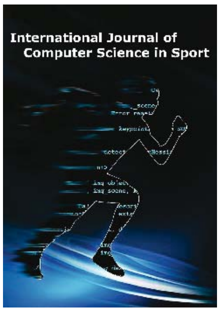
International Journal of Computer Science in Sport

Volume 18, Issue 3, 2019

Journal homepage: http://iacss.org/index.php?id=30

sciendo

DOI: 10.2478/ijcss-2019-0018



Assessing the causal impact of the 3-point per victory scoring system in the competitive balance of LaLiga

Soto-Valero, C[1], *Pic, M.*[2]

[1]*KTH Royal Institute of Technology, Sweden*

[2]*Department of Specific Didactics, Universidad de La Laguna, Spain*

Abstract

Competitive balance is a key concept in sport because it creates an uncertainty on the outcome that leads to increased interest and demand for these events. The Spanish Professional Football League (LaLiga) has been one of the top European leagues in the last decade, and it has given rise to a particular research interest regarding its characteristics and structure. Since season 1995/96, LaLiga changed the number of points given to the winning teams, by awarding three points per victory instead of two. In this paper, we assess the impact of such a change on the competitive balance of LaLiga. Our analysis focuses on teams with varying levels of performance and follows a three-step approach. First, we cluster the teams according to their historical performance using an adjusted measure based on their credible intervals of winning ratios. Second, we calculate Kendall's tau coefficient (according to our adjusted measure) in order to obtain the overall ranking turnover of teams between consecutive seasons. Third, we assess the causal impact of the adoption of the new scoring system, based on Kendall's tau coefficients, for each cluster of teams. Our results show that the overall competitive balance decreased after the adoption of the new scoring system. However, the impact was not the same for all teams, being more significant for top teams and less significant for bottom teams. Moreover, our predictions using adjusted ARIMA models indicate that this difference in the competitive balance will persist for future seasons.

KEYWORDS: FOOTBALL, COMPETITIVE BALANCE, CAUSAL INFERENCE, HIERARCHICAL K-MEANS, CLUSTERING, TIME SERIES FORECASTING

Introduction

In economics, competitive balance refers to the situation in which no business in a group of competing businesses has an unfair advantage over the others (Fort, 2007). The concept has been successfully applied to professional sport leagues, and has received considerable attention in the field of sport economics (Humphreys, 2002; Mourão & Teixeira, 2015). In this context, it is widely acknowledged that competitive balance has a direct impact on the demand for attending sporting events because the uncertainty of outcome directly instigates the fans' interest in the game (Forrest, Beaumont, Goddard, & Simmons, 2005; Haugen, 2008; Schmidt & Berri, 2001). From an economic point of view, the predictability of a match not only decreases the attractiveness of the match itself but also the profits obtained from the match. Numerous studies have proven that increasing competitive balance is in the best interest of sport leagues (Fort & Maxcy, 2003; Szymanski, 2001; Szymanski & Késenne, 2004).

Football is considered the most globally played sport. European football leagues attract the most attention nowadays due to their concentration of quality via talents acquisition and effective world-wide scouting (Littlewood, Mullen, & Richardson, 2011). In particular, the victories of the teams belonging to the Spanish Professional Football League (LaLiga) in the last international competitions (e.g. Champions League from 2016 to 2018, Euros 2008 and 2012, World Cup 2010) have motivated a special research interest regarding the competitive peculiarities of Spanish football.

One of the most controversial characteristics of the Spanish league is the imbalance between a few very successful teams and all the other teams. As a matter of fact, this excessive bi-polarization of actual victory chances reveals a clear gap in terms of competitive balance between the top teams historically, such as Real Madrid and Barcelona, and the other competing teams. As consequence, many argue that the Spanish league is one of the most unbalanced leagues in the world, and numerous studies have shown some reasons that fairly support such a claim (Vales-Vázquez, Casal-López, Gómez-Rodriguez, & Blanco-Pita, 2017). Among the factors that contribute to this phenomenon are the income disparity among the different clubs and administrations in the Spanish league, as well as the lack of competitiveness and support of some specific regions to their clubs. These factors were pointed out as being strong causes of this situation (Montes, Sala-Garrido, & Usai, 2014).

Until season 1994/95, teams in LaLiga received two points for a win, one point for a draw, and no points for a loss. However, since season 1995/96, the competition adopted the 3-point for a win scoring system. Although favoring the offensive game was one of the arguments used to justify such a change, it is still unclear to what extent this contributed to the increase or decrease of the competitive balance in the Spanish league (Kent, Caudill, & Mixon Jr, 2013). At the present time, the consequences of adopting that scoring system continue to be unclear (Moschini, 2010), and this is even more so if we take into account that the lack of competitive balance in LaLiga now seems to becoming much more obvious from one season to the next (Brocas & Carrillo, 2004; Groot, 2008).

In this paper, we investigate the impact of the change in the scoring system on the competitive balance of the Spanish league. We perform a large-scale empirical study on the competitive balance between-seasons, by using the match results of all teams that participated in LaLiga from season 1928/29 to 2016/17. We hypothesize that the competitiveness of the teams is divided into different levels, according to the characteristics and possibilities of various clusters of teams. In this context, it would be controversial to consider that the competitiveness of a regular league necessarily depends on the number of candidate teams that are able to take the first place in the league. A deeper look reveals that, while the teams in the middle of the table are highly competitive, maybe the teams at the bottom are even more competitive.

Following this idea, we propose an adjusted measure of historical performance based on the credible intervals of winning ratios to determine the level of performance of the teams and clustering them accordingly. Then, we apply two widely used measures of competitive balance between-seasons and a causal impact inference model in order to shed light on the effects of the 3-point scoring system on each cluster. In addition, we fit ARIMA models to predict the behavior of the competitiveness for future seasons, taking into account the performance of each cluster. In general, our results show that the competitive balance of LaLiga decreased after the adoption of the 3-point scoring system. However, such impact was not the same for each cluster of teams, being more evident in the group of top teams and less significant for the bottom teams. The result of our predictions until 2030 indicate that, if the adopted scoring system does not change, the current perceived difference of competitive balance will remain stable, or even increase, for future seasons.

In a previous work, Sánchez, Garcia-Calvo, Leo, Pollard, and Gómez (2009) show that the inclusion of the 3-point system reduced the home advantage of teams in LaLiga and the Spanish second division, as happened previously in the Premier League (Jacklin, 2005). However, unlike other related works in the literature that deal with the issue of analysing competitive balance without taking into account the different substructures and levels of performance of teams (Cardoso Marques, 2009; Easton & Rockerbie, 2005; Lopez, 2015), this paper aims to assess such differences by measuring the impact of a concrete phenomenon on it. Precisely, our analysis considers that identifying the competitive specificities of the teams according to their performance levels is essential to reveal the real competitiveness of the league as a whole. Therefore, the main contribution of this paper consists of assessing the competitive balance from distinct levels of performance using a novel Bayesian model of causal impact inference. To the best of our knowledge, no previous applications of Bayesian structural time series models for causal impact inference have been reported in the literature of sports science.

The remainder of this paper is structured as follows. Section 2 (Methods) describes the data used in this work as well as the proposed measure of historical performance for clustering the teams. This section also provides details about the measures of competitive balance applied to each cluster and offers a necessary introduction to causal impact inference. Section 3 (Results) presents our findings regarding the competitive balance of each individual group of Spanish teams, in comparison with the overall competitive balance of the league, and depict future predictions for the next decade. Section 4 (Discussion) offers a discussion of our results and Section 5 (Conclusion) summarizes our contributions and concludes the paper.

Methods

Data description

We analyze a data set with all games played in LaLiga between seasons 1928/1929 and 2016/2017. Data source was obtained from the official database of Juan Maria Alfaro[1]. The data includes the 62 teams that have participated in the league from the very beginning and the results (wins, losses or draws) of a total of 23564 matches.

Figure 1 shows, in chronological order, the name of the teams that won at least one title in LaLiga. Notice that the competition was suspended between 1937 and 1939 due to the Spanish Civil War. The columns show the teams that won titles in each decade and the numbers of

[1] http://tercertiempo.es

titles won are in brackets below the name of the teams. The colored arrows connect the teams that won titles in more than one decade.

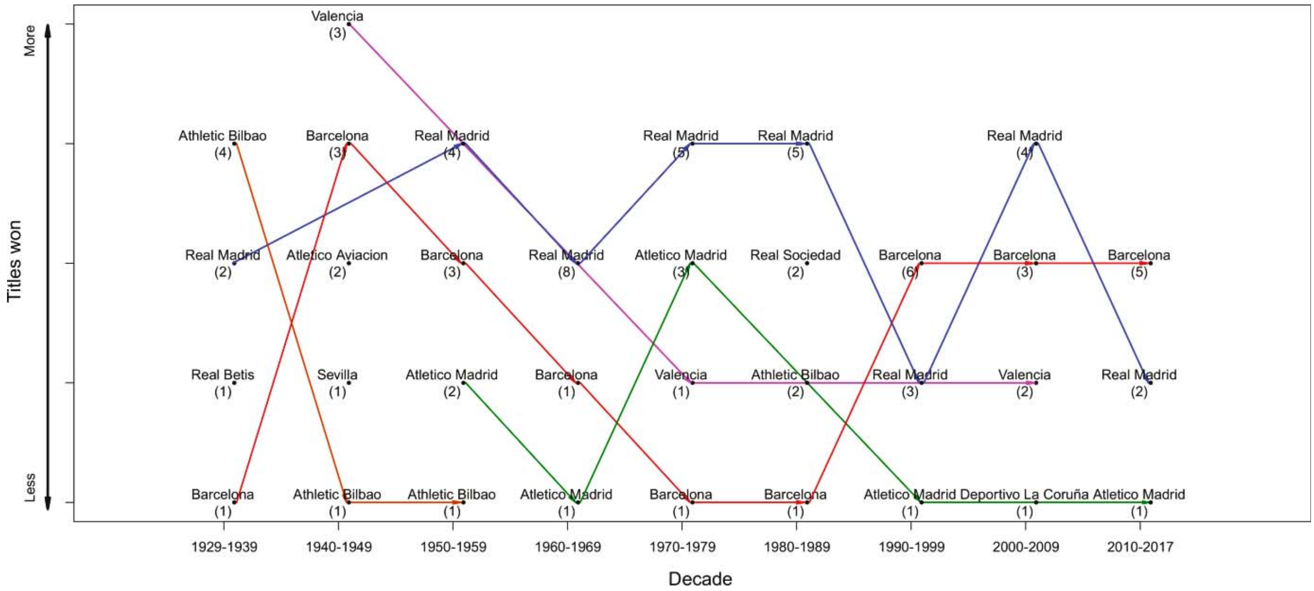


Figure 1: Teams that claimed titles in LaLiga, sorted top-down according to the number of titles gained during each decade. The number of titles is in brackets below the name of the team.

In the early years of LaLiga, the Athletic Bilbao team (later known as Athletic Club) claimed several titles but then Real Madrid dominated the championship from the 1950s right through to the 1980s. From the 1990s onwards, Barcelona and Real Madrid were both notable, though LaLiga also had some other champions, including Atlético Madrid, Valencia, and Deportivo La Coruña. In the 2010s, Atlético Madrid became increasingly strong, forming a trio alongside Real Madrid and Barcelona which occupied the podium places exclusively. In total, only 9 teams have been crowned champions in this tournament, with Real Madrid (33 titles) being at the top of the winners, followed by Barcelona (24 titles) and Atlético de Madrid (10 titles). The only three teams that have been present in all editions of the competition are Real Madrid, Barcelona and Athletic Club.

By observing this figure, it could be argued that the competitive balance between-seasons in LaLiga may be affected mainly due to the very low ranking mobility of the top teams throughout consecutive seasons. In this paper, we hypothesize that, beyond this fact, the implementation of the 3-point scoring system in season 1995/96 also had a significant impact on the competitive balance of LaLiga.

Data preprocessing

Adjusting win ratios

The Win Ratio (WR) is a standard measure of general performance in sports. The WR of a team is calculated as the number of victories divided by the total number of games played. This measure is one way to compare the records of two teams only in terms of the wins achieved, regardless of whether the victory awarded two or three points to the winner.

In general, the WR is a good descriptor of performance when enough data is provided. However, it lacks of accuracy as a predictor of future performance at the start of the season. For instance, consider the case of a team that loses its first game, at this moment it has no sense in predicting that this team will never get a win in the whole season because the knowledge about it in that moment is only based on little prior negative knowledge. We know that in football history most team have a WR between 0.3 and 0.5 in the season, with some rare exceptions on either side. Accordingly, we know that if a team loses a few games in a row at the start, it might indicate that it will end up a bit worse than average, but we know it probably will not deviate too much from that range. Moreover, the number of games played is a strong indicator for making good predictions regarding future performance.

In the particular case of LaLiga, Figure 2 shows that there are teams, such as Atlético Aviación (nowadays Atlético Madrid), that achieved a very high WR with less than 200 games played. This makes a general comparison unfair because we cannot say that Atlético Aviación is better than Atlético de Madrid just because it has a higher WR value. Thus, we need a better ranking approach that takes into account such differences.

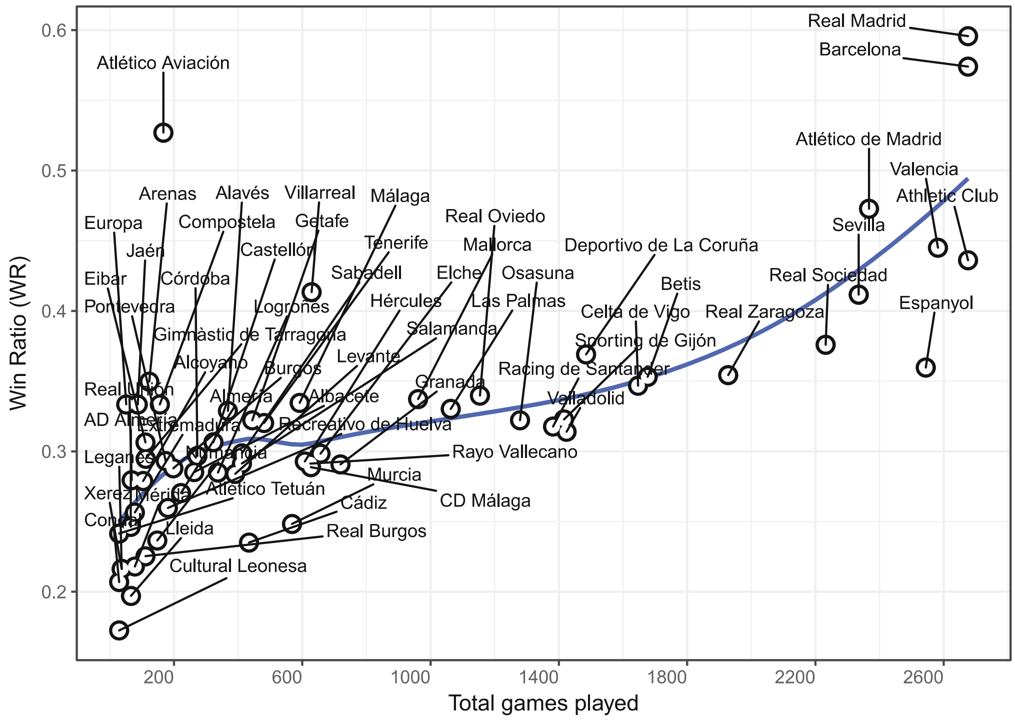


Figure 2: Historical performance of the teams in LaLiga according to WR measure of performance.

The results of a team in a season can be represented with a binomial distribution as a series of wins and losses (draws are not considered in this case). A natural way to represent its prior expectations is with the beta distribution. The domain of the beta distribution is $[0, 1]$ and the results are just like a probability estimation. This approach is called Beta-Binomial regression.

We fit a Beta-Binomial probability distribution $X \sim Beta(\alpha_0, \beta_0)$ to the data according to the WR of all teams. The model consists of adjusting the distribution of WR by taking into account the number of games played. This adjustment gives us a value of $\alpha_0 = 10.63$ and $\beta_0 = 23.01$. In order to rank each team according to Empirical Bayes (EB) estimates, we add α_0 to the number of wins, and $\alpha_0 + \beta_0$ to the total number of games played. Equation 1 shows the formula used to calculate the Win Ratio Adjusted (WRA) for the teams in Laliga.

$$WRA = \frac{W_t + \alpha_0}{W_t + L_t + D_t + \alpha_0 + \beta_0} = \frac{W_t + 10.63}{W_t + L_t + D_t + 33.74} \quad (1)$$

To illustrate this approach, Figure 3 shows the teams sorted in descending order according to the number of games they played, which indicates how much information we have about them. The vertical dashed red line represents the quotient $\alpha_0/(\alpha_0 + \beta_0)$, which is the mean WRA

throughout history (based on our Beta fit) for all teams. The credible interval indicates that 95% of the posterior distribution lies within the confidence region. Notice that once there is enough information, the credible intervals and confidence intervals are nearly identical, but in the cases of the teams Atlético Tetuán (7/29) and Condan (6/29) teams, the credible interval is much narrower. Hence, the EB estimations bring us information about the individual performance of teams based on knowledge obtained from the full data set.

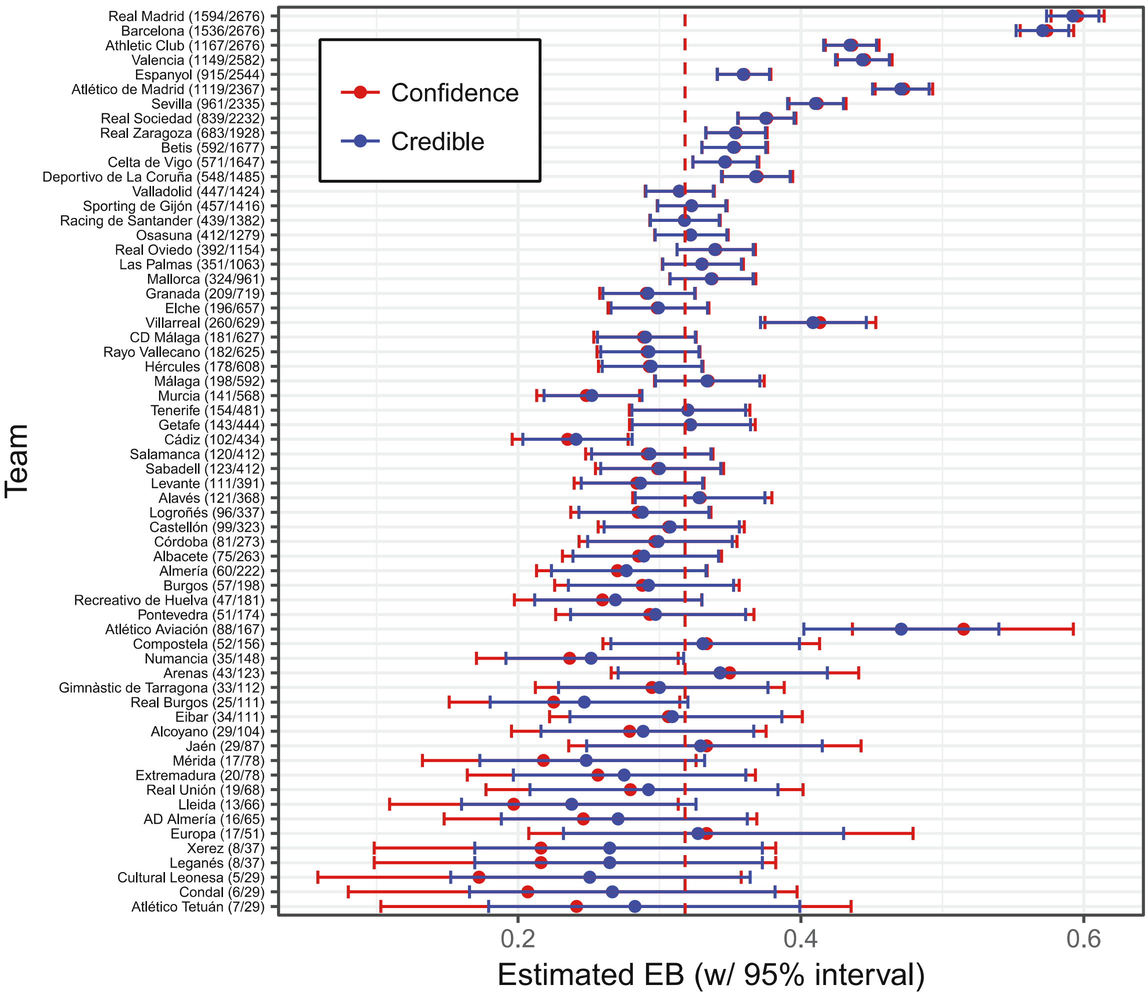


Figure 3: Credible and confidence intervals of winning ratios for all teams in LaLiga according to historical Empirical Bayes estimates (within 95% of the credible interval).

Clustering teams

We cluster the teams according to their performance over time using the WRA measure described in the previous section. In order to perform the clustering, the input data is modelled in the form of a $n \times m$ matrix. Each row of the matrix ($n = 62$) represents a team to be clustered and we have as many columns as the number of seasons to be included for our comparison purposes ($m = 88$). For instance, consider the Atlético Aviación team, it is modelled as a row in the matrix with WRA values in the columns that represent the seasons between 1940 and 1946 (the period played by this team). In the case of years during which the teams did not play, their WRA values in the matrix are set to be NA.

In order to perform the clustering, we use the hierarchical k-means clustering algorithm, which is a hybrid approach that combines the advantage of high accuracy of hierarchical clustering and fast convergence of k-means (Lu, Tang, Tang, & Yang, 2008). The algorithm requires the definition of a dissimilarity measure to be used in order to partition the data space into teams with similar performances. This leads to the construction of a dissimilarity matrix D, whose entries $d_{i,j}$ with $i, j = 1, \dots, n$ are the dissimilarities between any pair of teams i and j. The calculation of $d_{i,j}$ can be obtained in different ways, depending on the kind of data to be clustered. In this case, we use the Euclidean Distance (ED) as the dissimilarity measure, which is the standard choice when the descriptors lie in the same domain and have the same scale of measurement. Equation 2 shows this distance, calculated for two teams i and j.

$$ED(i,j) = \sqrt{\sum_{t=1}^{n} (WRA_i - WRA_j)^2} \tag{2}$$

Figure 4 shows the result obtained after the application of the clustering method described above. Three major clusters were identified in the data. These clusters are consistently associated with different levels of performance relating to the teams, which we named as top teams, middle teams and bottom teams. To facilitate the analysis, in the rest of this work we link each team with one of these categories. This distinction allow us to assess the competitive balance of each cluster separately.

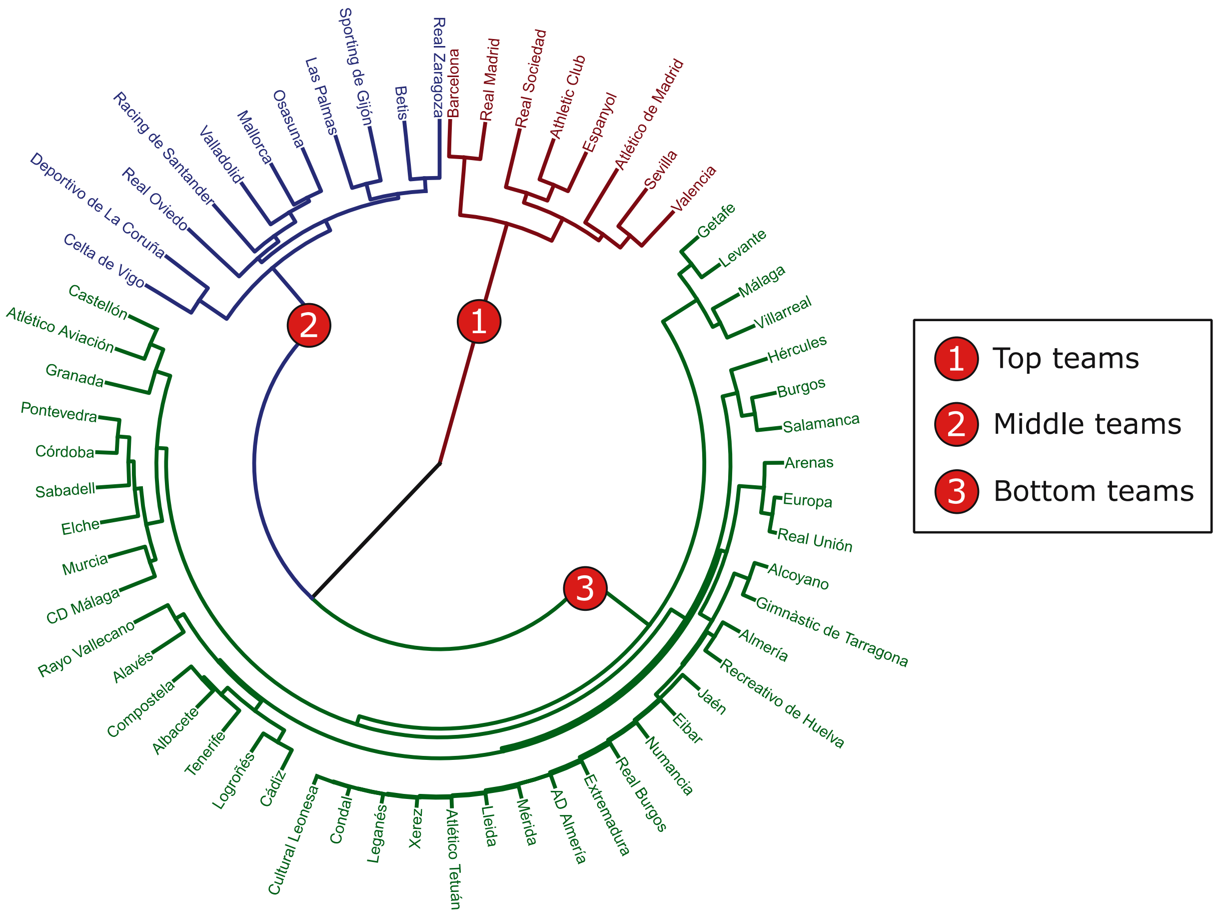


Figure 4: Hierarchical k-means clustering results for the football teams in LaLiga according to their historical WRA performance scores.

Measures of competitive balance

Relative standard deviation of season performance

The Actual Standard Deviation (ASD) of the teams' WRA is a natural measure of competitive balance for a season. It is calculated as follows:

$$ASD = \sqrt{\sum_{t=1}^{n} \frac{(WRA_t - 0.5)^2}{n}} \tag{3}$$

where WRA_t is the Win Ratio Adjusted of team t and n is the total number of teams that competed in the season. A smaller value of ASD indicates a more equal competition. Therefore, an increase in ASD has negative impacts in the competitive balance.

When comparing ASD values, either for the same league over time or across different leagues, the number of teams and games played are typically not constant. Other things being equal, the ASD tends to decrease as the number of games played increases because the prior assumptions lie within a more credible interval and there is likely to be less random noise in the final outcomes. Hence, it is useful to compare the ASD with a Idealized Standard Deviation (ISD) benchmark, corresponding to a prior representation of a perfectly balanced league in which each team has the same probability of winning each game.

The ISD can be derived as the standard deviation of a binomially distributed random variable with a (constant) probability of success of 0.5 across independent trials. It is calculated as follows:

$$ISD = \frac{0.5}{\sqrt{G}} \tag{4}$$

where G is the number of games played in each season.

The Relative Standard Deviation (RSD) is the ratio between the ASD and the ISD. It is calculated as follows:

$$RSD = \frac{ASD}{ISD} \tag{5}$$

Therefore, as the WRA increases, any reduction in the WRA will be compared against the reduced value of the ISD benchmark. The ideal perfectly balanced league has an RSD value which equals zero. The higher the ratio, the more the actual spread diverges from the ideal one and hence the lower the competitive balance is (and greater is the gap that separates the teams in terms of their perceived performance).

Kendall's tau correlation coefficient of performance between-seasons

The Kendall's tau coefficient (τ) is a common metric used to calculate the overall ranking turnover of the teams in a football league (Groot, 2008). In this work, we compute τ to calculate the differences in the competitive balance between-seasons, which allows us to quantify the relative quality of teams across (of at least two) consecutive seasons (Manasis & Ntzoufras, 2014). In the particular case of European football, this approach is more informative than calculating the balance within a single season because it takes into account the ranking mobility (Leeds & von Allmen, 2009).

The definition of τ is based on the number of transpositions s required to transform a particular rank order into another specific order. In this work, the rank order of each season is determined by using the WRA value of each team in the season. In essence, the rank s is compared with the maximum number of possible transpositions$-s_{max}$ of the ranking, which is equal to $\frac{n(n-1)}{2}$. The formula of the τ index is given by

$$\tau = 1 - \frac{2s}{s_{max}} = 1 - \frac{4s}{n(n-1)} \tag{6}$$

and its original definition lies in the interval from -1 to 1. In this work, we denote by τ' the following rescaled modification:

$$\tau' = 1 - \frac{1+\tau}{2} = 1 - \frac{2s}{n(n-1)} \tag{7}$$

This rescaled version of the index lies in the interval from 0 to 1, which corresponds to the cases of a perfectly balanced and completely unbalanced league, respectively.

Causal impact inference

Causal impact analysis is a statistical approach used to estimate the causal effect of a designed intervention on an outcome variable over time (Camillo & D'Attoma, 2010). It has been applied mostly in economics to infer the causal impact of a market intervention, such as a new product launch or the onset of an advertising campaign (Hoover, 2012; Lewis, Rao, & Reiley; Rubin & Waterman, 2006). In such context, the causal impact of an intervention can be defined as the difference between the observed value of the response and the (unobserved) value that would have been obtained if the intervention had not taken place. That is, the quantification of the estimated effect of the intervention on the intervened (Chan, Ge, Gershony, Hesterberg, & Lambert). In the case of time series, the causal impact is the difference between the observed series and a counterfactual series which represents the values that would have been observed in the absence of the intervention (Winship, 2007).

Brodersen Kh (2015) proposed a Bayesian Structural Time Series ($BSTS$) model to perform causal impact inference when there is no straightforward candidate series for the counterfactual. This model constructs an adequate synthetic control time series using information provided by the behaviour of other time series that were not affected by the intervention (Abadie, Diamond, & Hainmueller, 2010). The counterfactual is constructed using the predictions of the response series after the intervention, based on the combination of a set of candidate controls via Bayesian averaging (Hoeting, Madigan, Raftery, & Volinsky, 1999).

By using this model, we focus on measuring the impact of the adoption of the 3-point scoring system in the competitive balance of LaLiga. Answering this type of question can be difficult when a randomized experiment is not available. In this setting the response variable is a time series, so the causal effect of interest is the difference between the observed series and the series that would have been observed if the 3-point per victory had never happened. Given a response time series (i.e. Kendall's tau coefficients of competitive balance between sequential seasons for a group of teams) and a set of control time series which were not directly affected by the intervention (i.e. the Adjusted Win Ratios of teams across seasons), a $BSTS$ model is constructed. This model is then used to try and predict the counterfactual, i.e. how the competitive balance would have evolved after the intervention if the intervention had never occurred.

We use the R-package CausalImpact[2], which provides an open-sourced implementation of causal impact inference model, this package is originally developed by Google (Brodersen Kh, 2015). Our $BSTS$ model uses static regression coefficients, which assumes that the linear competitive balance between the control and the counterfactual expected competitiveness remains constant even after the adoption of the new scoring systems in the 1995/96 season. Given a response time series y composed by Kendall's tau correlation coefficients of competitive balance between sequential seasons and a control time series of adjusted win ratios of teams across seasons, the $BSTS$ model can be defined as

$$y_t = \mu_t + Z_t + \varepsilon_t \tag{8}$$

where the first component μ_t is the value of the trend at time t. The $\mu_{t+1} = \mu_t + \delta_t + \eta_{\mu,t}$ captures the random walk $\delta_{t+1} = \delta_t + \eta_{\delta,t}$ and local linear trends make the model sensible to unobserved sources of variability in the time series. The regression component $Z_t = \beta' x$ captures the linear relation between the control series and the objective series. The error term $\varepsilon_t \sim \mathcal{N}(0, \sigma_t^2)$ follows an independent Gaussian distribution, as well as $\mu_{\mu,t} \sim \mathcal{N}(0, \sigma_{\mu,t}^2)$ and $\mu_{\delta,t} \sim \mathcal{N}(0, \sigma_{\delta,t}^2)$.

By using the combined information of the target time series y_t and the controls, we estimate the posterior distribution of the counterfactual time series y_t' The difference $y_t' - y_t$ yields a semiparametric Bayesian distribution which can be used to obtain credible intervals about how the competitive balance would have evolved after the adoption of the 3-point scoring system if it had never been adopted in LaLiga.

Results

Historical trend of competitive balance per season

We use RSD of the WRA to compare the actual standard deviation of win ratios for each season with respect to the idealized scenario in which each team that participated in the season has an equal chance of winning each game. This measure of seasonal performance takes into account both the season length and the number of teams, thus facilitating a comparison of the competitive balance over time in the league (Fort & Lee, 2007).

Figure 5 shows the seasonal competitive balance of LaLiga from 1929 to 2017. The fitted linear regression line reveals a historical increase in the RSD values (and hence a decrease in the competitive balance of the league). We also observe a slight decrease in RSD values during the recent seasons, which suggests an increase in the within-season competitive balance after the change in the scoring system. Nevertheless, the historical trend illustrates the general concern regarding the domination of a few teams that win much more games than the rest and manifest a tendency to remain in the first places season after season.

[2] http://google.github.io/CausalImpact

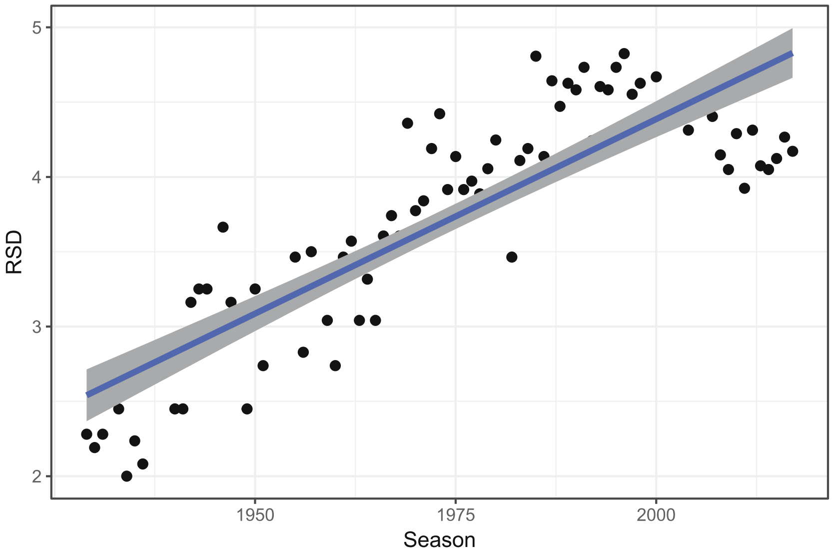
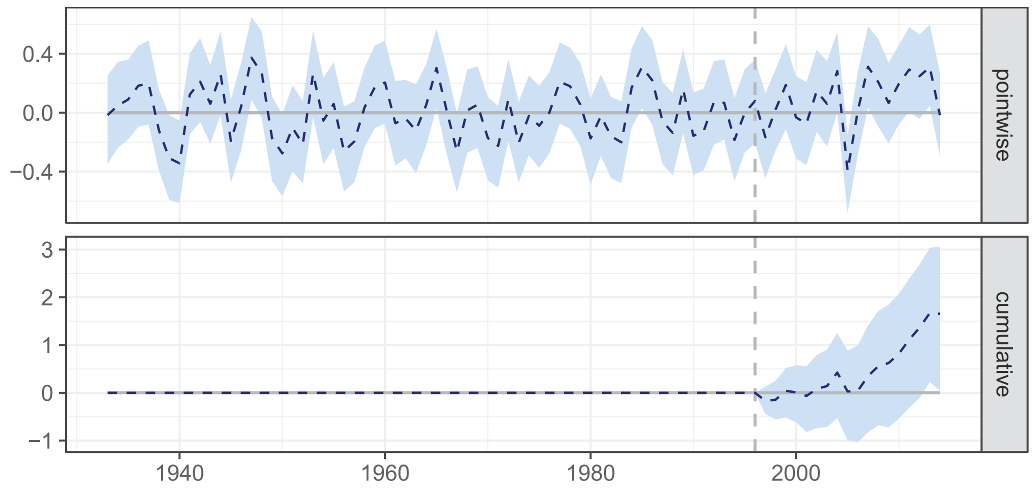


Figure 5: Overall trend in the between-seasons competitive balance of LaLiga using the RSD measure.

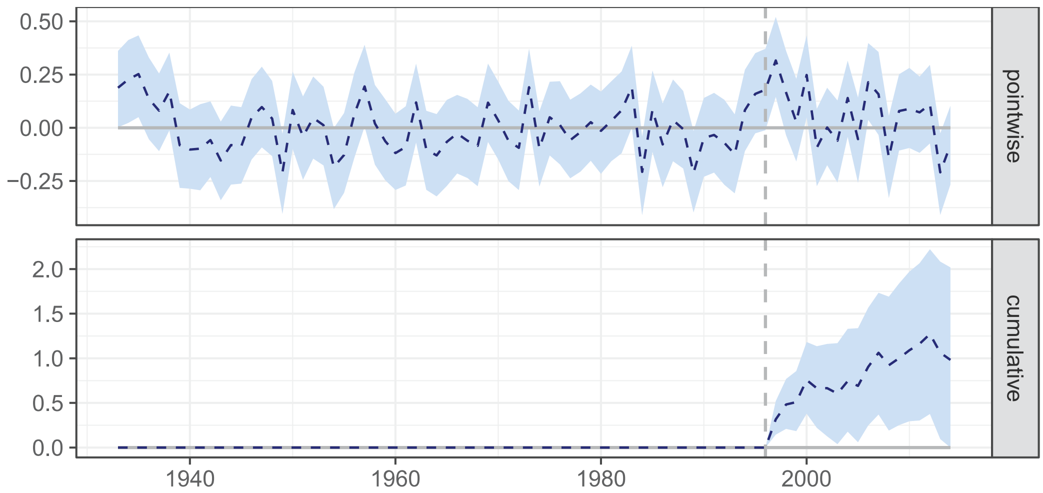
However, the RSD values just give a partial view of the evolution of competitive balance (Owen, 2010). They do not account for championship concentration or persistence of performance of individual teams over successive seasons (see Figure 1). For this aim we need to use between-seasons competitive balance measures, which consider the ranking mobility of teams with dissimilar performance. In the next section, we evaluate the competitive balance for each cluster of teams using the τ' coefficient described in previously. Furthermore, the causal impact of the new scoring system in each particular cluster is assessed.

Causal impact of the new scoring system

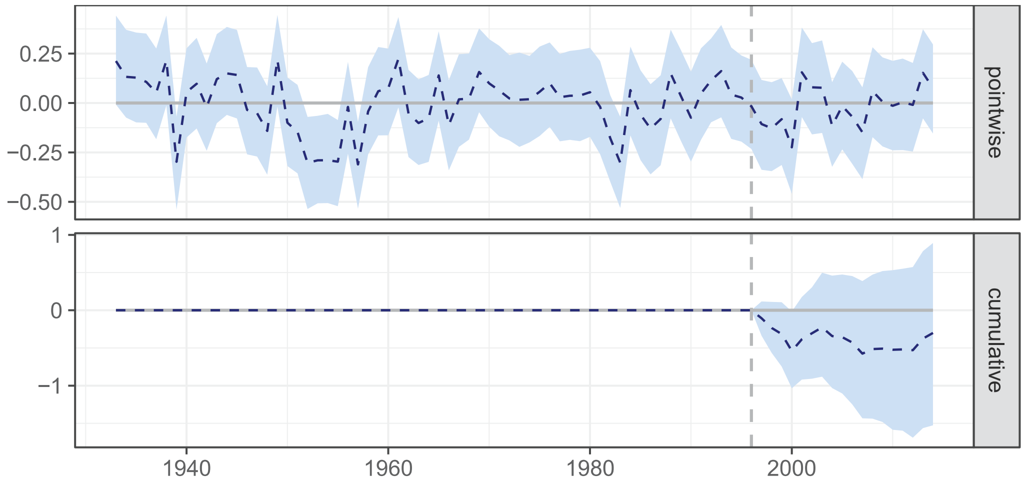
Figure 6 shows the causal impact result obtained for each cluster. The intervention point, indicated by a vertical dashed red line, separates the periods before and after the inclusion of the 3-point scoring system (pre-intervention and post-intervention periods). In all the panels, 95% prediction confidence intervals are indicated with a shaded light blue color. The point-wise panels shows the difference between observed competitive balance and counterfactual predictions. This is the point-wise causal effect, as estimated by the BSTS model. The cumulative panel adds up the contributions from the point-wise panel, representing the cumulative effect (positive, negative, or null) of the intervention.



(a) Top teams.



(b) Middle teams.



(c) Bottom teams.

Figure 6: Causal impact of the three-point scoring system on the competitive balance for the three clusters of teams in LaLiga.

In addition, Table 1 presents summary statistics regarding the causal impact on each cluster. It shows estimated values of the causal effect of the intervention according to the actual prediction for the response variable. The last row shows the statistical test of significance of the results. We provide additional details on these results in the following subsections.

Table 1: Summary results of the causal impact inference model applied to the clusters of teams in LaLiga.

	Cluster #1		**Cluster #2**		**Cluster #3**	
	Mean	Cumulative	Mean	Cumulative	Mean	Cumulative
Actual prediction	0.65 (0.53)	13.64 (11.06)	0.67 (0.62)	14.06 (12.93)	0.76 (0.77)	16.06 (16.20)
Absolute effect	0.12	2.57	0.054	1.136	-0.0067	-0.1414
Relative effect	23%	23%	8.80%	8.80%	-0.87	-0.87%
p-value	p = 0.002		p = 0.001		p = 0.418	

Detailed impact on the top teams

In the case of teams in Cluster #1 (top teams) during the post-intervention period, the response variable had an average value throughout time of approximately 0.65. In the absence of an intervention, we would have expected an average response of 0.53. The 95% interval of this counterfactual prediction is [0.45, 0.60]. Subtracting this prediction from the observed response yields an estimate of the causal effect the intervention had on the response variable. This effect is 0.12 with a 95% interval of [0.046, 0.20].

Adding up the individual data points during the post-intervention period, the response variable had an overall value of 13.64. By contrast, had the intervention not taken place, we would have expected a sum of 11.06. The 95% interval of this prediction is [9.38, 12.67].

The above results are given in terms of absolute numbers. In relative terms, the response variable showed an increase of +23%. The 95% interval of this percentage is [+9%, +38%]. This means that the positive effect observed during the intervention period is statistically significant and unlikely to be due to random fluctuations. It should be noted, however, that the question of whether this increase also bears substantive significance can only be answered by comparing the absolute effect (0.12) with the original goal of the underlying intervention. The probability of obtaining this effect by chance is very small (Bayesian one-sided tail-area probability p = 0.002). This means the causal effect can be considered statistically significant.

Detailed impact on the middle teams

In the case of teams in Cluster #2 (middle teams) during the post-intervention period, the response variable had an average value of approximately 0.67. In the absence of an intervention, we would have expected an average response of 0.62. The 95% interval of this counterfactual prediction is [0.57, 0.67]. Subtracting this prediction from the observed response yields an estimate of the causal effect the intervention had on the response variable. This effect is 0.054 with a 95% interval of [0.0021, 0.10].

When adding up the individual data points during the post-intervention period, the response variable had an overall value of 14.06. By contrast, had the intervention not taken place, we would have expected a sum of 12.93. The 95% interval of this prediction is [11.87, 14.02].

The above results are given in terms of absolute numbers. In relative terms, the response variable showed an increase of +9%. The 95% interval of this percentage is [+0%, +17%]. This means that the positive effect observed during the intervention period is statistically significant and unlikely to be due to random fluctuations. It should be noted, however, that the question of whether this increase also bears substantive significance can only be answered by comparing the absolute effect (0.054) with the original goal of the underlying intervention. The probability of obtaining this effect by chance is very small (Bayesian one-sided tail-area probability p = 0.001). This means the causal effect can be considered statistically significant.

Detailed impact on the bottom teams

In the case of teams in Cluster #3 (bottom teams) during the post-intervention period, the response variable had an average value of app approximately 0.76. In the absence of an intervention, we would have expected an average response of 0.77. The 95% interval of this counterfactual prediction is [0.70, 0.84]. Subtracting this prediction from the observed response yields an estimate of the causal effect the intervention had on the response variable. This effect is -0.0067 with a 95% interval of [-0.073, 0.060].

When adding up the individual data points during the post-intervention period, the response variable had an overall value of 16.06. Had the intervention not taken place, we would have expected a sum of 16.20. The 95% interval of this prediction is [14.80, 17.60].

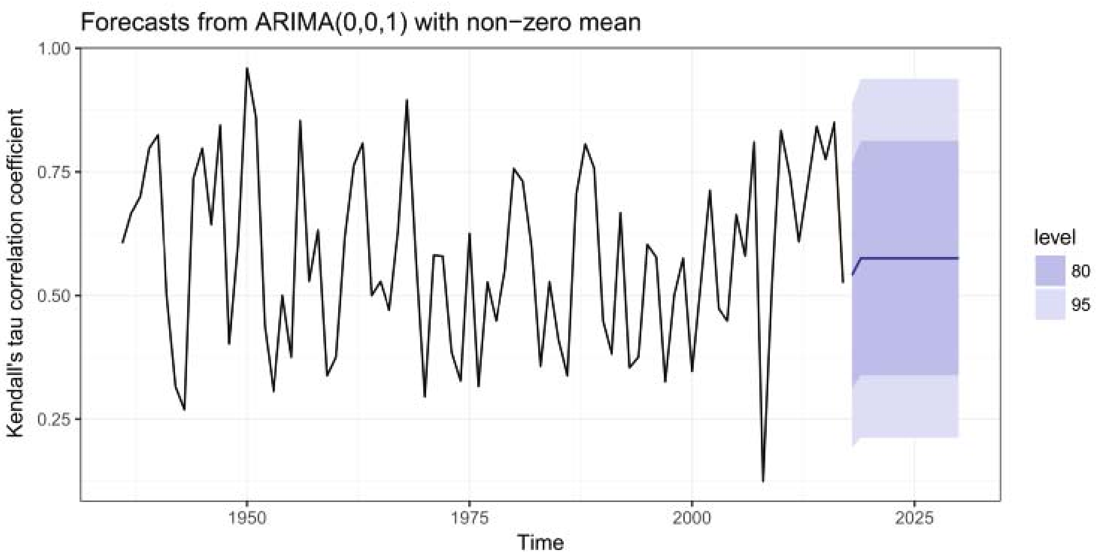
The above results are given in terms of absolute numbers. In relative terms, the response variable showed a decrease of -1%. The 95% interval of this percentage is [-10%, +8%]. This means that, although it may look as though the intervention has exerted a negative effect on the response variable when considering the intervention period as a whole, this effect is not statistically significant, and so cannot be meaningfully interpreted. The apparent effect could be the result of random fluctuations that are unrelated to the intervention. The probability of obtaining this effect by chance is p = 0.439. This means the effect may be spurious and would generally not be considered statistically significant.

Future predictions of competitive balance

We are also interested in performing an estimation of the future competitive balance in LaLiga. For this aim, we use the R-package forecast[3], which provides a set of methods and tools for displaying and analysing univariate time series forecasts, including exponential smoothing, via state space models and automatic autoregressive moving average (ARIMA) modelling.

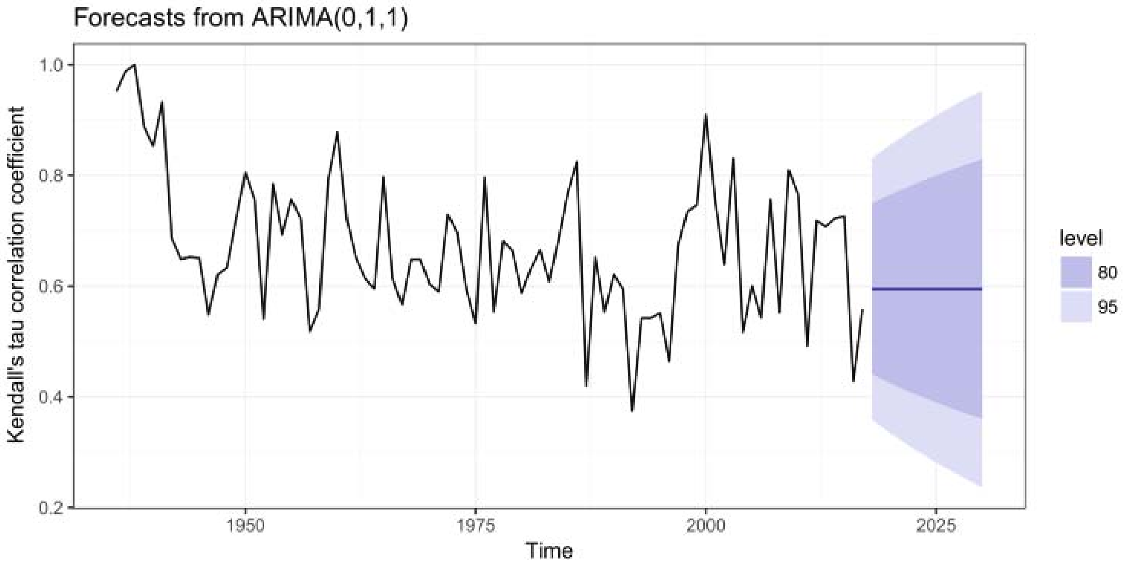
We obtained optimal ARIMA models using the automatic forecasting algorithm proposed by Hyndman and Khandakar (2008), which combines unit root tests, minimization of the AIC and MLE to obtain an adjusted ARIMA model. Figure 7 shows the time series prediction results from 2017/18 to 2029/30 for each cluster of teams according to their Kendall's tau correlation coefficient of performance between-seasons. The shaded areas correspond to 80% and 95% prediction intervals.

The predictive results suggest a slight decrease in the competitive balance of the top teams for the future. The prediction for the middle teams is that their competitive balance will remain stable during this period. On the other hand, the predictive results for the bottom teams indicate that we can expect a slight increase in their competitive balance.

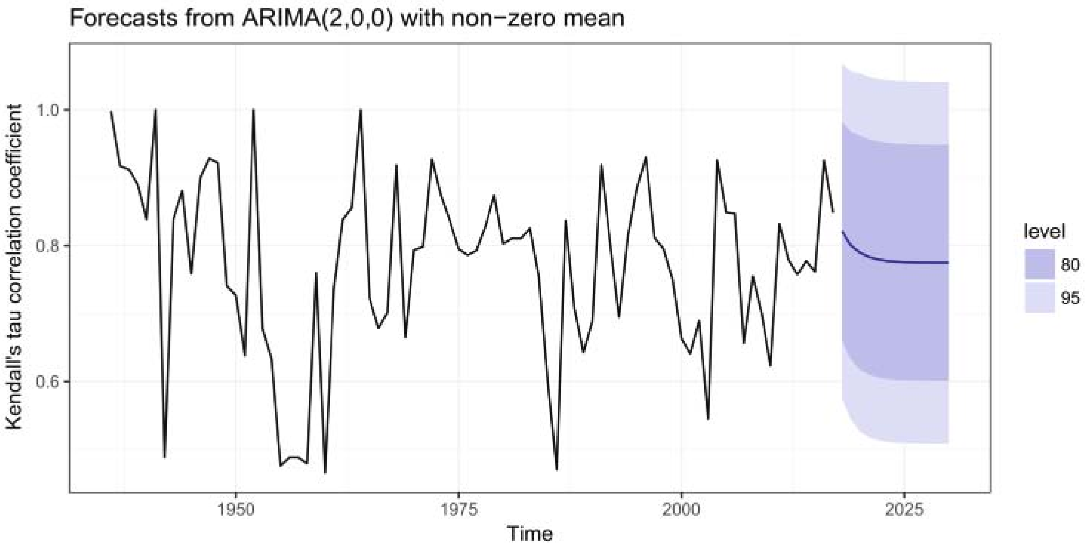


(a) Top teams.

[3] https://github.com/robjhyndman/forecast



(b) Middle teams.



(c) Bottom teams.

Figure 7: Predicted results of the competitive balance until 2030 for the three clusters of teams in LaLiga.

Discussion

Many different indices have been introduced to measure competitive balance in sport. For instance, the Herfindahl-Hirschman index (Owen, Ryan, & Weatherston, 2007), the standard deviation of performance (Criado, Garcia, Pedroche, & Romance, 2013) or the Kendall's tau correlation coefficient between seasons (Manasis & Ntzoufras, 2014), just to mention a few. However, as with many other measures adapted from elementary econometric concepts, competitiveness is a latent variable that cannot be measured directly, and as such its proper quantification still remains an important issue in the domain of sport analytics (Trandel & Maxcy, 2011). In general, there are two major approaches adopted in the literature for measuring competitive balance: within-season and between-seasons. The first one considers the relative quality of teams during a particular season, while the latter has a larger scope and refers to the perceived competitiveness of teams throughout several consecutive seasons.

In this work, we proposed an adjusted performance measure based on the EB estimates of a Beta-Binomial probability distribution of win ratios. This measure allows us to make a fairer comparison of the performance of the teams across a larger time period, by calculating credible intervals that take prior information into account (e.g. the number of games played by the teams). The measure, called Win Ratios Adjusted (WRA), can also be used in any other sport to produce better comparisons among the teams. For example, in the case of searching for more accurate historical equivalences of teams' performance, or as a dissimilarity measure for ranking tasks.

Our clustering results show that the Spanish Professional Football League presents different levels of competitiveness. On the one hand there is a small group of very successful teams, led by Real Madrid and Barcelona, with a high rate of victories per match, while on the other hand there is a larger group of teams that compete almost on equal terms. This result is not surprising, as it is in line with existing studies that analyse the particular characteristics of LaLiga (Vales-Vázquez et al., 2017).

We believe that this division of teams in terms of their levels of quality and possibilities has a high impact on how the competitive balance can be measured and interpreted. Consequently, it could be unfair to claim that the Spanish league is the most unbalanced league in Europe just because there are a few teams with a very high budget (e.g. Real Madrid and Barcelona with 507 and 500 million EUR for the 2017/18 season, respectively) which can afford to hire the best players. Out of this group of top teams, we found a group of middle teams with a similar historical performance, as well as a group of bottom teams with a lower performance. These distinct levels of quality produce different levels of competitiveness that must be taken into account during the global analysis of the competitive balance of the league. In this context, the Bosman ruling in 1995 had a profound effect on the transfers of footballers within the European Union, contributing to increase the existing gap between rich and modest teams.

It is still expected that a more equal league in the economic sense should be a more competitively equal league, and consequently a better and more attractive sporting spectacle (Szymanski, 2003). Previous economic studies have analysed the negative effects of imbalance in LaLiga (Montes et al., 2014; Mourão & Teixeira, 2015; Ramchandani, 2012). An interesting starting point for future work is to investigate to what extent the concentration of quality in only a few teams, as in the case of LaLiga, contributes to increase or decrease the quality of the football spectacle as a whole by considering a larger set of quantitative and qualitative variables.

Our results confirm that the promotion-relegation rule, which is a structural characteristic of the European football, greatly contributes towards a more competitive league and, as a consequence, it is a beneficial mechanism for improving competitiveness in LaLiga (Manasis & Ntzoufras, 2014). This rule produces a ranking mobility that is a strong incentive to compete, even for the bottom teams. We also want to point out that, from the fans' point of view, the foundation of some rivalries should be searched in sources other than performance (Abrevaya, 2004; Karanfil, 2017).

With respect to the repercussion of the implementation of the 3-point scoring system, our causal impact analysis shows that it had a different influence on each group of teams. The increase of rewarding points for the winner clearly decreases the competitive balance in the group of top teams, separating the elite teams much more from the rest. With regards to the middle teams, which is the largest group according to our clustering results, we found that the new rule also had a negative impact on the competitive balance. In contrast, the new scoring system had a low impact for the bottom teams, being mostly beneficial for them. This result could be interpreted in different ways, as one can assume that granting more points for a win gives an advantage to the better teams, but also decreases the interest in the results of the league (Dilger & Geyer, 2009). What is clear is that the inclusion of the 3-point scoring system reduced the competitiveness of LaLiga in general, as occurred in other European leagues (Hon & Parinduri, 2016). A similar impact could also be hypothesised in the Spanish Second Division. In our opinion, further research should be done on the scoring systems, in order to establish a more effective threshold between the concentration of quality and the competitiveness of the league (Banerjee, Swinnen, & Weersink, 2007).

Our analysis presents some limitations. As with other approaches referring to causal inference, valid conclusions require strong assumptions. For instance, we used the time series of adjusted win ratios of teams across seasons as controls, and other measures of competitive balance between-seasons could also be used as counterfactual. On the other hand, it might be too simplistic to conclude that the only cause of the decrease in competitiveness in LaLiga was the implementation of the 3-point system. It could be a rather much more complex multi-factorial phenomenon with other fundamental explanations. Thus, the lack of consideration of more diverse exogenous and endogenous factors is a current limitation of this study. Furthermore, the number of existing measures to evaluate the competitive balance, both static and dynamic, is very large and the small number of them we have used could raise some doubts about the generalization of our results (Zheng, Oh, Kim, Dickson, & De Bosscher, 2017). However, our approach could be extended with a larger set of measures in order to provide more information to the inferring model. In addition, although our causal impact analysis is for football, this methodology can be applied to other sports as well. Furthermore, in order to promote open science and facilitating the replication of our results, we made all our data analysis publicly available online[4].

Conclusion

In this paper, we investigated how the change in the scoring system impacted the competitive balance of the Spanish Professional Football League. We observed a low ranking mobility of the top teams from season 1928/29 to season 2016/17, which suggests the existence of a clear gap of performance among different groups of teams. Such differences were modelled using an adjusted measure of win ratios (WRA) based on Empirical Bayes estimates of prior expectations, which allow us to get more concise knowledge about the individual historical performance of each team. The teams were clustered according to their WRA values using a hierarchical k-means clustering method. The clustering model revealed three major clusters, which we identified as top teams, middle teams and bottom teams.

Two measures of competitive balance were taken into account in order to characterize the competitive balance of each cluster of teams: the relative standard deviation of season's performance (RSD) and the Kendall's tau correlation coefficient of performance between-seasons. Overall, the RSD values confirm the historical decrease in competitive balance observed in previous research works. Beyond this observation, we conducted a causal impact analysis of competitive balance between-seasons which showed that the disparity is not the same for all clusters of teams.

In summary, the results of our causal impact inference analysis show that the overall competitive balance of LaLiga decreased after the adoption of the 3-point scoring system. This makes sense due to the possible increase in points of the best teams, which augments the gap between top teams and all the other teams. Cluster #1 (top teams) showed a significant

[4] https://github.com/cesarsotovalero/ijcss-comp-balance-laliga

decrease in the competitive balance after the adoption of the 3-point scoring system. Cluster #2 (middle teams) showed a significant decrease in the competitive balance after the adoption of the 3-point scoring system. Cluster #3 (bottom teams) showed a low increase in the competitive balance after the adoption of the 3-point scoring system. In addition, ARIMA models were utilized to predict the competitive balance of each cluster until 2030. Our results indicated that the current gap in the competitive balance among the different groups of teams in LaLiga will continue during future seasons.

References

Abadie, A., Diamond, A., & Hainmueller, J. (2010). Synthetic control methods for comparative case studies: Estimating the effect of California's tobacco control program. *Journal of the American Statistical Association, 105*(490), 493-505.

Abrevaya, J. (2004). Fit to be tied: The incentive effects of overtime rules in professional hockey. *Journal of Sports Economics, 5*(3), 292-306.

Banerjee, A. N., Swinnen, J. F., & Weersink, A. (2007). Skating on thin ice: rule changes and team strategies in the NHL. *Canadian Journal of Economics/Revue canadienne d'économique, 40*(2), 493-514.

Brocas, I., & Carrillo, J. D. (2004). Do the "three-point victory" and "golden goal" rules make soccer more exciting? *Journal of Sports Economics, 5*(2), 169-185.

Brodersen Kh, G. F. K. J. R. N. S. S. L. (2015). Inferring causal impact using Bayesian structural time-series models. *Annals of Applied Statistics, 9*(1), 247-274.

Camillo, F., & D'Attoma, I. (2010). A new data mining approach to estimate causal effects of policy interventions. *Expert Systems with Applications, 37*(1), 171-181.

Cardoso Marques, A. n. (2009). Estimating Quality in Home Advantage and Competitive Balance in the Portuguese Football League. *Journal of Quantitative Analysis in Sports, 5*(2).

Chan, D., Ge, R., Gershony, O., Hesterberg, T., & Lambert, D. (2010). *Evaluating online ad campaigns in a pipeline: causal models at scale.* Paper presented at the Proceedings of the 16th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining.

Criado, R., Garcia, E., Pedroche, F., & Romance, M. (2013). A new method for comparing rankings through complex networks: Model and analysis of competitiveness of major European soccer leagues. *Chaos: An Interdisciplinary Journal of Nonlinear Science, 23*(4), 043114-043114.

Dilger, A., & Geyer, H. (2009). Are three points for a win really better than two? A comparison of German soccer league and cup games. *Journal of Sports Economics, 10*(3), 305-318.

Easton, S. T., & Rockerbie, D. W. (2005). Overtime! Rules and incentives in the national hockey league. *Journal of Sports Economics, 6*(2), 178-202.

Forrest, D., Beaumont, J., Goddard, J., & Simmons, R. (2005). Home advantage and the debate about competitive balance in professional sports leagues. *Journal of Sports Sciences, 23*(4), 439-445.

Fort, R. (2007). Comments on``Measuring Parity". *Journal of Sports Economics, 8*(6), 642-651.

Fort, R., & Lee, Y. H. (2007). Structural change, competitive balance, and the rest of the major leagues. *Economic Inquiry, 45*(3), 519-532.

Fort, R., & Maxcy, J. (2003). Competitive balance in sports leagues: An introduction. *Journal of Sports Economics, 4*(2), 154-160.

Groot, L. (2008). *Economics, Uncertainty and European Football: Trends in Competitive Balance*: Northampton: Edward Elgar Publishing Limited.

Haugen, K. K. (2008). Point score systems and competitive imbalance in professional soccer. *Journal of Sports Economics, 9*(2), 191-210.

Hoeting, J. A., Madigan, D., Raftery, A. E., & Volinsky, C. T. (1999). Bayesian model averaging: a tutorial. *Statistical Science*, 382-401.

Hon, L. Y., & Parinduri, R. A. (2016). Does the three-point rule make soccer more exciting? Evidence from a regression discontinuity design. *Journal of Sports Economics, 17*(4), 377-395.

Hoover, K. D. (2012). Economic theory and causal inference. *Philosophy of Economics, 13*, 89-113.

Humphreys, B. R. (2002). Alternative measures of competitive balance in sports leagues. *Journal of Sports Economics, 3*(2), 133-148.

Hyndman, R. J., & Khandakar, Y. (2008). Automatic time series forecasting: the forecast package for R. *Journal of Statistical Software, 26*(3), 1-22.

Jacklin, P. B. (2005). Temporal changes in home advantage in English football since the Second World War: What explains improved away performance? *Journal of Sports Sciences, 23*(7), 669-679.

Karanfil, F. (2017). An empirical analysis of European football rivalries based on on-field performances. *Sport Management Review, 20*(5), 468-482.

Kent, R. A., Caudill, S. B., & Mixon Jr, F. G. (2013). Rules changes and competitive balance in European professional soccer: Evidence from an event study approach. *Applied Economics Letters, 20*(11), 1109-1112.

Leeds, M. A., & von Allmen, P. (2009). *The Economic of Sports*: The Pearson Series in Economics.

Lewis, R. A., Rao, J. M., & Reiley, D. H. (2011). *Here, there, and everywhere: correlated online behaviors can lead to overestimates of the effects of advertising.* Paper presented at the Proceedings of the 20th International Conference on World Wide Web.

Littlewood, M., Mullen, C., & Richardson, D. (2011). Football labour migration: An examination of the player recruitment strategies of the 'big five'European football leagues 2004–5 to 2008–9. *Soccer & Society, 12*(6), 788-805.

Lopez, M. J. (2015). Inefficiencies in the National Hockey League points system and the teams that take advantage. *Journal of Sports Economics, 16*(4), 410-424.

Lu, J. F., Tang, J. B., Tang, Z. M., & Yang, J. Y. (2008). Hierarchical initialization approach for K-Means clustering. *Pattern Recognition Letters, 29*(6), 787-795.

Manasis, V., & Ntzoufras, I. (2014). Between-seasons competitive balance in European football: review of existing and development of specially designed indices. *Journal of Quantitative Analysis in Sports, 10*(2), 139-152.

Montes, F., Sala-Garrido, R., & Usai, A. (2014). The lack of balance in the Spanish first division football league. *European Sport Management Quarterly, 14*(3), 282-298.

Moschini, G. (2010). Incentives and Outcomes in a Strategic Setting: The 3-Points-for-a-Win System in Soccer. *Economic Inquiry, 48*(1), 65-79.

Mourão, P. R., & Teixeira, J. S. (2015). Gini playing soccer. *Applied Economics, 47*(49), 5229-5246.

Owen, P. D. (2010). Limitations of the relative standard deviation of win percentages for measuring competitive balance in sports leagues. *Economics Letters, 109*(1), 38-41.

Owen, P. D., Ryan, M., & Weatherston, C. R. (2007). Measuring competitive balance in professional team sports using the Herfindahl-Hirschman index. *Review of Industrial Organization, 31*(4), 289-302.

Ramchandani, G. (2012). Competitiveness of the English Premier League (1992-2010) and ten European football leagues (2010). *International Journal of Performance Analysis in Sport, 12*(2), 346-360.

Rubin, D. B., & Waterman, R. P. (2006). Estimating the causal effects of marketing interventions using propensity score methodology. *Statistical Science*, 206-222.

Sánchez, P. A., Garcia-Calvo, T., Leo, F. M., Pollard, R., & Gómez, M. A. (2009). An analysis of home advantage in the top two Spanish professional football leagues. *Perceptual and Motor Skills, 108*(3), 789-797.

Schmidt, M. B., & Berri, D. J. (2001). Competitive balance and attendance: The case of Major League Baseball. *Journal of Sports Economics, 2*(2), 145-167.

Szymanski, S. (2001). Income inequality, competitive balance and the attractiveness of team sports: Some evidence and a natural experiment from English soccer. *The Economic Journal, 111*(469), F69-F84.

Szymanski, S. (2003). The economic design of sporting contests. *Journal of Economic Literature, 41*(4), 1137-1187.

Szymanski, S., & Késenne, S. (2004). Competitive balance and gate revenue sharing in team sports. *The Journal of Industrial Economics, 52*(1), 165-177.

Trandel, G. A., & Maxcy, J. G. (2011). Adjusting winning-percentage standard deviations and a measure of competitive balance for home advantage. *Journal of Quantitative Analysis in Sports, 7*(1).

Vales-Vázquez, A., Casal-López, C., Gómez-Rodriguez, P., & Blanco-Pita, H. (2017). A comparison of competitive profiles across the Spanish football leagues. *International Journal of Computer Science in Sport, 16*(3), 207-220.

Winship, C. (2007). *Counterfactuals and causal inference: Methods and principles for social research*: Cambridge University Press.

Zheng, J., Oh, T., Kim, S., Dickson, G., & De Bosscher, V. (2017). Competitive balance trends in elite table tennis: the Olympic Games and World Championships 1988-2016. *Journal of Sports Sciences*, 1-9.