Estimation of final standings in football competitions with premature ending: the case of COVID-19

P. Gorgi[a,b], S. J. Koopman[a,b,c], and R. Lit[a]

[a]Vrije Universiteit Amsterdam, The Netherlands

[b]Tinbergen Institute, The Netherlands

[c]Aarhus University, Denmark

October 5, 2020

Abstract

We study an alternative approach to determine the final league table in football competitions with a premature ending. For several countries, a premature ending of the 2019/2020 football season has occurred due to the COVID-19 pandemic. We propose a model-based method as a possible alternative to the use of the incomplete standings to determine the final table. This method measures the performance of the teams in the matches of the season that have been played and predicts the remaining non-played matches through a paired-comparison model. The main advantage of the method compared to the incomplete standings is that it takes account of the bias in the performance measure due to the schedule of the matches in a season. Therefore, the resulting ranking of the teams based on our proposed method can be regarded as more fair in this respect. A forecasting study based on historical data of seven of the main European competitions is used to validate the method. The empirical results suggest that the model-based approach produces more accurate predictions of the true final standings than those based on the incomplete standings.

Key words: Bivariate Poisson, COVID-19, paired-comparison models, sport statistics.

1 Introduction

The socio-economic impact of COVID-19 on our society has been overwhelming. Sport events have not been an exception and they have been heavily affected by the COVID-19 pandemic. Major sport events such as the Olympic Games, the UEFA European Championship and the Tour de France have been postponed or canceled. Several ongoing sport competitions, including some of the main European football competitions, have experienced a premature ending. The premature ending of a football competition raises the issue of how to settle its final table. This has created some public debate in the media (newspapers, radio and TV) and on social media. The final standings of a competition are important to determine promotions and relegations and to select the teams that take part in international competitions, for the next season. A possible solution to determine the final standings is to consider the position of the teams in the table at the time when the competition has prematurely ended, which we refer to as the incomplete standings. This has been the mainstream approach for several football leagues. For instance, in the French Ligue 1, the average number of points per match at the time of the stop has been used to determine the final table. Similarly, in the Dutch Eredivisie, the incomplete standings has been used as the final table to determine the teams that qualify for European competitions.

The idea of using the incomplete table to determine the final standings can be justified as a ranking of the teams based on their merit in the games that have been played before the premature ending. In principle, this should reflect the expected performance in the remaining games and deliver a fair ranking of the football teams. However, the incomplete standings suffer some drawbacks for this purpose. The strength of the opposing teams in the remaining part of the competition may differ among teams. One team may have already played against all the strong teams in the competition while another team may still need to face the stronger opponents. This creates an imbalance and favors teams that have strong opponents left in the games after the premature ending. Another shortcoming of using the incomplete standings concerns home and away games. The presence of a significant home ground advantage in football matches is well documented in the literature; see, for example, Pollard (2008). Different teams can have a different number of home and away games left to be played and this would favor teams that have already played more home games before the premature ending of the competition.

In this paper, we consider an alternative model-based approach that takes into account the strength of the opposing teams as well as the home ground advantage. We measure the performance of the teams by means of a statistical paired-comparison model. The Bradley-Terry model is a traditional example of a paired-comparison model; see Bradley and Terry (1952) and, for a review, Cattelan (2012). The outcome of a match is taken as a paired-comparison observation for the two teams that are involved in the match. In a paired-comparison model, the strength level of a team is measured relative to the strength level of the opposing team. In our analysis we determine the final ranking based on the performance of the teams in earlier match results in the season, those taking place before the premature ending. Once the strength levels of the teams have been measured for each possible match, the model can be employed to predict the expected number of points in each of the remaining games. Finally, the expected number of points at the end of the season can be used to rank the teams and obtain the final standings. We adopt a paired-comparison model that is closely related to the model of Maher (1982), which has become a standard approach in the literature to describe the outcome of football matches.

To validate whether the proposed model-based approach provides a better measure of the performance of the teams, when compared to the incomplete standings, we conduct an empirical study based on a longitudinal dataset that consists of 25 seasons of seven major European football leagues (England, Spain, Germany, Italy, France, Portugal and Netherlands). For each season, we artificially stop each competition at some selected point and obtain the final table from the incomplete standings and from the model-based approach. We treat the standings that are predicted from the two methods as forecasts of the actual final table, and we measure the accuracy of the forecasts using Kendall's tau correlation. Finally, we construct a longitudinal test to verify whether the difference in the accuracy of the forecasts from the two methods is statistically significant. The results show evidence that the model-based standings better reflect the true final standings. These findings suggest that a model-based method is more accurate in determining the final table, in case the season has ended prematurely. The model-based method is more fair as it discounts the effect of the schedule of the matches and it is a more accurate in providing the final standing of the season.

The remainder of the paper is organized as follows. Section 2 discusses the details of our model-based approach to determine the final standings using a statistical analysis. Section 3 presents the testing methodology to compare the forecasting performance of the model-based approach with the incomplete standings in forecasting the true final table. Section 4 reports the empirical results. Section 5 concludes.

2 A statistical method to estimate the final standings

When the incomplete standings are used to set the final ranking, it does not take into account the bias introduced by the schedule of the games and the different skill levels of the opposing teams, in the remaining part of the season (with games that will never be played due to the premature ending). For example, assume there are a few games left to the end of the season and consider two teams, team A and team B, with the same number of points. Team A has already played the strongest opponents while instead team B still needs to face some strong opposing teams in the last games. In such a case, it may be desirable to take into account that team A has shown a better performance than team B: although the two teams have the same number of points, team A has faced the stronger teams. Furthermore, team A has "easier" matches left to be played and therefore team A can be expected to collect more points. Another drawback of taking the incomplete table as the final ranking is that home and away games are not accounted for. Teams can have a different number of games to be played at home. It is well-known that the team playing at home obtains a higher likelihood of gaining more points from the game; see the discussions in Pollard (2008) and Buraimo et al. (2012).

The approach we propose is designed to take these factors into consideration. We measure the performance of the teams in the season by means of a paired-comparison model as in Maher (1982). The performance of the teams is obtained only using the outcomes of matches that have already been played in the same season. On the basis of this measured performance, we determine the final standings using the model-implied expected number of points in the remaining games. We should emphasize that our analysis is based on a paired-comparison model that has been widely used to model and predict football matches by including regression variables as well as time-variation in the strength of the teams. In particular, the paired-comparison model for outcomes of football matches of Maher (1982) is adopted in many studies, including Dixon and Coles (1997), Goddard (2005), Karlis and Ntzoufras (2009), Hvattum and Arntzen (2010), Rue and Salvesen (2000) and Koopman and Lit (2015, 2019). However, the purpose of the current study is not to construct a predictive model using all the available data but to provide a fair measure of the performance of the teams in the current season to determine the final table. We provide a detailed description of the approach in the remainder of this section.

We denote the outcome of a football match between the home team i and the away team j as a pair of counts (X_i, Y_j) for $i, j \in \{1, \ldots, n\}$, $i \neq j$, where X_i is the number of goals scored by the home team i, Y_j is the number of goals scored by the away team j, and n indicates the number of teams in the competition. We describe the match outcome by means of a bivariate Poisson paired-comparison model

$$(X_i, Y_j) \sim \mathcal{BP}(\lambda_{ij}^x, \lambda_{ij}^y, \gamma),$$

where $\mathcal{BP}(\lambda_{ij}^x, \lambda_{ij}^y, \gamma)$ denotes a bivariate Poisson distribution with intensity λ_{ij}^x for the home team count, intensity λ_{ij}^y for the away team count, and coefficient γ for the dependence between the two counts. The probability mass function (pmf) of the bivariate Poisson $\mathcal{BP}(\lambda_{ij}^x, \lambda_{ij}^y, \gamma)$ is given by

$$\mathbb{P}(X_i = x, Y_j = y) = \frac{\lambda_{ij}^x \, \lambda_{ij}^y}{x! \, y! \, e^{\lambda_{ij}^x + \lambda_{ij}^y + \gamma}} \sum_{k=0}^{\max\{x,y\}} \binom{x}{k} \binom{y}{k} k! \left(\frac{\gamma}{\lambda_{ij}^x \lambda_{ij}^y} \right)^k, \tag{1}$$

for $x, y \in \{0, 1, \ldots, \infty\}$. We refer to Johnson et al. (1997) for a review of the bivariate Poisson distribution and to Karlis and Ntzoufras (2003) for its original application to sport matches. The intensities λ_{ij}^x and λ_{ij}^y determine the difference in expected goals between the home and away teams. The intensities are specified as

$$\lambda_{ij}^{x} = \exp(\delta + \alpha_i + \beta_j), \quad \text{and} \quad \lambda_{ij}^{y} = \exp(\alpha_j + \beta_i),$$

where α_k represents the attacking ability of team k, $k = i, j$, β_k represents the defending ability of team k, $k = i, j$, and δ is the home ground advantage. The specification for the intensities originates from Maher (1982). It accounts for the different strength level of the teams (α_k, β_k), $k = 1, \dots, n$, as well as the home ground advantage δ to determine the probability distribution of the match outcome.

In most applications of the bivariate Poisson model for sports data, the objective is to specify α_k and β_k to best predict the outcomes of future matches. For instance, the model can extended with other covariates that may explain the strength levels of the team. However, in our current study the objective is to determine a final ranking that reflects the performance of the teams in the matches that have been played earlier in the season. We achieve this by estimating the parameters of the model $(\alpha_1, \dots, \alpha_n, \beta_1, \dots, \beta_n, \delta, \gamma)$ using the method of maximum likelihood (ML) and only based on the data of the current season. In this way, the estimated intensities only reflect the performance of the teams in the current season. The estimation is subject to the standard restriction $\sum_{i=1}^{n} \alpha_i = 0$ for the purpose of identification of the parameters, since only the differences between the attack and defense strengths are identified and not their overall level. Once the parameters have been estimated, we can obtain the expected number of points of each team in the remaining games and construct the final table by using the model-implied expected number of points at the end of the season. In particular, first we calculate the winning probability of the home team p_h, the winning probability of the away team p_a, and the probability that the match ends with a draw p_d as follows

$$p_h = \sum_{y=0}^{\infty} \sum_{x=y+1}^{\infty} \mathbb{P}(X_i = x, Y_j = y),$$

$$p_a = \sum_{x=0}^{\infty} \sum_{y=x+1}^{\infty} \mathbb{P}(X_i = x, Y_j = y),$$

and

$$p_d = \sum_{z=0}^{\infty} \mathbb{P}(X_i = z, Y_j = z),$$

where the expression of the pmf is given in equation (1). Based on these probabilities, we calculate the expected number of points of the home and away teams. We consider the system of assigning 3 points to the winning team, 0 to the losing team, and 1 point to each of the teams if the game ends with a draw. This system is the standard in most football competitions. The expected number of points of the home team ep_h and the one of the away team ep_a are

$$ep_h = 3p_h + p_d, \quad \text{and} \quad ep_a = 3p_a + p_d.$$

The final table is obtained by summing up the expected number of points of each team in the remaining games and adding these expected points to the points of the incomplete table.

3 Testing the relative accuracy of estimated final standings

Once the final table is estimated using the method of the previous section, we need to verify whether the model-based approach provides a better estimate of the final ranking of the teams, compared to taking the incomplete table as the final ranking. We emphasize that both the model-based estimate of the final standings and the incomplete standings are based on the same data set consisting of all match results in the season, before the premature ending of the competition. Both estimates can be interpreted as forecasts of the true final table. We can use historical data from several football competitions to test whether the model-based approach performs better than the incomplete standings in terms of forecast accuracy. Next we present our proposed testing methodology to compare the two forecasts of the final table.

We can assess the accuracy of a final table forecast using the tau correlation coefficient of Kendall (1938) between the predicted and the actual table rankings. The Kendall tau statistic is a correlation measure between two rankings. The higher the correlation, the more similar the rankings. Assume that we have historical dataset of K competitions observed over T seasons. We can select a premature stopping time for the seasons. For instance, we can stop the seasons after a certain percentage of games have been played in the competition. Then, for competition i, with $i = 1, \ldots, K$, and season t, with $t = 1, \ldots, T$, we obtain two Kendall tau correlations: (i) between the model-based prediction and the true final table $\tau^m_{i,t}$, and (ii) between the incomplete table prediction and the true final table $\tau^c_{i,t}$. We define the difference between the two correlation coefficients $\tau^m_{i,t}$ and $\tau^c_{i,t}$ as $d_{i,t} = \tau^m_{i,t} - \tau^c_{i,t}$. Next, we can formally test the hypothesis that the expected difference $\mu_i = \mathrm{E}(d_{i,t})$ is different from zero. When $\mu_i > 0$, the model-based prediction is more accurate than the incomplete table prediction in forecasting the true final table. Hence, we consider the following test hypothesis

$$H_0 : \; \mu_i = 0 \quad \text{against} \quad H_1 : \; \mu_i \neq 0. \tag{2}$$

The test statistic is given by

$$s_i = \sqrt{T} \, \frac{\bar{d}_i}{\widehat{\sigma}_i},$$

where T is the sample size (number of seasons) and

$$\bar{d}_i = \frac{1}{T} \sum_{t=1}^{T} d_{i,t}, \qquad \widehat{\sigma}_i = \sqrt{\frac{1}{T} \sum_{t=1}^{T} (d_{i,t} - \bar{d}_i)^2}.$$

Under standard conditions, the statistic s_i has an approximate standard normal distribution. We note that the standard deviation $\hat{\sigma}_i$ is obtained under the assumption that the difference of Kendall tau statistics $d_{i,t}$ is not autocorrelated over time t. This assumption may be relaxed using a robust estimate of the standard error. However, we have not found any evidence of serial autocorrelation in the series $d_{i,1}, \ldots, d_{i,T}$, for any $i = 1, \ldots, K$, in our empirical study of Section 4.

The hypothesis (2) can be considered to verify whether the model-based predictions are more accurate for a given competition i, for $i = 1, \ldots, K$. The drawback of testing for each competition separately is that the test statistic will have low power to reject the null hypothesis since the number of seasons T is relatively low. We therefore also consider a longitudinal test to pool information from multiple competitions together. A longitudinal test can be based on the hypothesis

$$H_0: \frac{1}{K}\sum_{i=1}^{K}\mu_i = 0 \quad \text{against} \quad H_1: \frac{1}{K}\sum_{i=1}^{K}\mu_i \neq 0. \tag{3}$$

If we assume that the expected difference in the Kendall tau coefficients is the same for all competitions $\mu_i = \mu$, for $i = 1, \ldots, K$, the test hypothesis in (3) reduces to $H_0 : \mu = 0$ against $H_1 : \mu \neq 0$. We retain the general form of the test without assuming $\mu_i = \mu$ to avoid unnecessary assumptions. The corresponding test statistic is given by

$$s = \sqrt{KT}\,\frac{\bar{d}}{\widehat{\sigma}},$$

where

$$\bar{d} = \frac{1}{K}\sum_{i=1}^{K}\bar{d}_i, \qquad \widehat{\sigma} = \sqrt{\frac{1}{K}\sum_{i=1}^{K}\hat{\sigma}_i^2}.$$

Under standard regularity conditions, the test statistic s has an approximate standard normal distribution.

4 Empirical evidence for seven European football leagues

In our empirical study, we consider six football competitions with the highest score in the 2019/2020 UEFA ranking system. These competitions are the football leagues of England, Spain, Germany, Italy, France and Portugal. More information about the UEFA ranking system can be found on the official UEFA website: `https://www.uefa.com/`. The dataset used in the study is obtained from the website `http://www.football-data.co.uk/`. We have selected the highest six countries in the 2019/2020 UEFA ranking. The 7th ranked country is Russia. We do not have the historical data available for the Russian football competition. Instead, we have included the Netherlands since their data is available, but also since it is one of the countries where the 2019/2020 football season was stopped prematuraly due to the COVID-19 restrictions imposed by the government. Within this group of seven countries in our empirical study, the two competitions of France and the Netherlands have stopped prematurely and the incomplete standings are used to determine the final table. The empirical study is based on historical data from the season 1994/1995 to the season 2018/2019 for the group of seven countries. Hence, we have 25 seasons for each competition. All computations and analyses are done by the software package *Time Series Lab - Sports Statistics Edition* of Lit (2020) which is freely available at `https://timeserieslab.com`. A step-wise procedure to replicate the results in Tables 2 and 3 can be found in the Online Appendix.

4.1 Testing the superior precision of the model-based approach

We apply the testing methodology described in Section 3. We consider several premature ending times of the seasons to see how the stopping point affects the results. In particular, we report the test results for the following percentages of games that have already been played before the stop of the season occurs: 50%, 60%, 70%, 80%, 90% and 95%. Table 1 reports the test statistics for each competition as well as the test statistic of the longitudinal test that includes all the competitions. If we focus on the test for each individual competition in (2), we can see that most of the test statistics are not significant at a significance level of 0.05 or 0.1. This result is not surprising since, the sample size T is very small, $T = 25$. Therefore, the test statistics are highly affected by sampling uncertainty and the power of the test is low. Overall, we see that most of the test statistics are positive. This suggests that the forecast of the end of the season standings obtained using the model-based approach may be more accurate than the incomplete table.

We now focus on the results of the longitudinal test in (3) that includes all seven competitions. For this test, the actual sample size is larger since the information from the cross sectional dimension is also exploited. The total number of observations used for the test is $KT = 175$, where $K = 7$ and $T = 25$. Therefore, this test is expected to have more power. From the results, we can see that the tests statistic is positive for all the completion levels of the competitions. Furthermore, the test is significant at 1% level for 80% and 90% completion of the season, it is significant at 5% level for 70% completion of the season, and it is significant at 10% level for 60% completion of the season. Instead, the results are not significant for 50% and 95% completion of the season. These findings are not surprising. When 50% of the season is completed, the teams are facing all opponents in the remaining games of the season. Therefore, the model-based approach will tend to produce similar results as the incomplete table. Differences can be due to the fact that the model-based approach measures the skills of the teams based on goals that are scored and conceded, instead, the incomplete table only accounts for the points, irrespective of the number of goals. When the completion level of the season is 95%, we also do not expect major differences in performance between the methods. In this case, there are only two games left to be played and therefore changes in the final table are less likely to occur. Overall, we can conclude that there is statistical evidence that the model-based approach produces better predictions of the true final standings.

Table 1: Test statistics for the seven competitions for different completion levels of the seasons. The last row reports the statistics for the longitudinal test that includes all seven competitions. The reported significance of the test is indicated by * (10% level), ** (5% level), and *** (1% level).

	50%	60%	70%	80%	90%	95%
England	3.14 ***	2.71 ***	-0.21	3.44 ***	2.53 **	0.60
Spain	0.53	1.35	0.94	1.03	1.24	-0.05
Germany	-1.10	0.65	0.91	1.29	-1.14	-0.70
Italy	1.99 **	2.57 **	1.68 *	1.11	1.88 *	2.09 **
France	-0.01	0.04	2.00 **	1.34	1.20	0.45
Portugal	-0.96	-1.79 *	0.83	-0.05	1.20	1.37
Netherlands	-0.32	0.34	-1.17	0.36	1.60	0.13
All countries	**0.63**	**1.83** *	**2.05** **	**3.23** ***	**3.10** ***	**1.25**

4.2 The premature endings of the French and Dutch competitions

Finally, we apply the method to the 2019/2020 season of the French Ligue 1 and the Dutch Eredivisie. Both these competitions were stopped and the final standings were settled using the incomplete table. For the French competition, the average point per match was used to determine the final standings. Obviously, if the teams have played the same number of games, then the average point per match and the incomplete standings give the same ranking of the teams. For the Dutch competition, the incomplete table was used to determine the final standings and select the teams entering the European competitions in the season 2020/2021. Tables 2 and 3 report the incomplete table and the model-based table for the French Ligue 1 and the Dutch Eredivisie, respectively.

Table 2: Incomplete table and model-based table of the 2019/2020 season of the French Ligue 1.

	Incomplete Listings				**Model-based Listings**		
	Team	Points	Matches		Team	Points	Matches
1	Paris SG	68	27	1	Paris SG	94.27	38
2	Marseille	56	28	2	Marseille	71.25	38
3	Rennes	50	28	3	Rennes	66.68	38
4	Lille	49	28	4	Lille	65.16	38
5	Reims	41	28	5	Lyon	57.96	38
6	Nice	41	28	6	Reims	55.05	38
7	Lyon	40	28	7	Nice	54.36	38
8	Montpellier	40	28	8	Monaco	54.22	38
9	Monaco	40	28	9	Montpellier	53.80	38
10	Angers	39	28	10	Bordeaux	53.80	38
11	Strasbourg	38	27	11	Strasbourg	52.71	38
12	Bordeaux	37	28	12	Nantes	51.14	38
13	Nantes	37	28	13	Angers	50.72	38
14	Brest	34	28	14	Brest	45.69	38
15	Metz	34	28	15	Metz	45.13	38
16	Dijon	30	28	16	Dijon	41.40	38
17	St Etienne	30	28	17	St Etienne	40.20	38
18	Nimes	27	28	18	Nimes	36.90	38
19	Amiens	23	28	19	Amiens	33.70	38
20	Toulouse	13	28	20	Toulouse	20.50	38

We learn from Table 2 that the most relevant difference in the French rankings of the Ligue 1 teams concerns Lyon. In the model-based standings, Lyon is ranked 5th instead of 7th in the incomplete table. Lyon is ranked ahead of Reims and Nice in the model-based standings, which is not the case in the incomplete standings. The first 6 positions of the French Ligue 1 are important to enter the European competition and the 5th position gives access to the UEFA Europa League. The schedule of the remaining games reveal that Lyon has already played the strongest team, Paris SG. Instead, both Reims and Nice have not faced Paris SG. Furthermore, Lyon has six home matches left compared to the five of Reims and the four of Nice. This may well explain the difference between the two rankings presented in Table 2

In case of the Dutch Eredivisie, the main difference between the two rankings in Table 3 is that PSV Eindhoven is in the 3rd position in the model-based standings, instead of

Table 3: Incomplete table and model-based table of the 2019/2020 season of the Dutch Eredivisie.

	Incomplete Listings				**Model-based Listings**		
	Team	Points	Matches		Team	Points	Matches
1	Ajax	56	25	1	Ajax	76.99	34
2	AZ Alkmaar	56	25	2	AZ Alkmaar	75.66	34
3	Feyenoord	50	25	3	PSV	65.77	34
4	PSV	49	26	4	Feyenoord	64.94	34
5	Willem II	44	26	5	Willem II	56.96	34
6	Utrecht	41	25	6	Utrecht	54.62	34
7	Vitesse	41	26	7	Vitesse	51.66	34
8	Heracles	36	26	8	Heracles	46.33	34
9	Groningen	35	26	9	Groningen	46.20	34
10	Heerenveen	33	26	10	Heerenveen	43.51	34
11	Sparta	33	26	11	Sparta	43.39	34
12	FC Emmen	32	26	12	FC Emmen	39.58	34
13	VVV Venlo	28	26	13	Twente	37.16	34
14	Twente	27	26	14	Zwolle	37.03	34
15	Zwolle	26	26	15	VVV Venlo	33.77	34
16	For Sittard	26	26	16	For Sittard	32.02	34
17	Den Haag	19	26	17	Den Haag	27.42	34
18	Waalwijk	15	26	18	Waalwijk	20.94	34

4th in the incomplete standings. We notice that PSV overtakes Feyenoord in the model-based standings despite having played one more match. Feyenoord has played 25 games and PSV 26 before the stop of the competition occurred. When taking a closer look to the schedule of the non-played matches, a similar situation as described for the French Ligue 1 occurs. PSV has already played against Ajax, which is the strongest team in the competition, instead Feyenoord has not played Ajax. Furthermore, PSV plays five of the remaining matches at home while Feyenoord only plays four matches at home.

5 Conclusion

We have presented and discussed a model-based approach to determine the final standings of football competitions with a premature ending. The key advantage of the model-based approach is that it accounts for the schedule of the matches when measuring the performance of the teams and when predicting the final table. The empirical study to seven of the main European competitions indicates that the model-based approach tends to deliver a final table that is closer to the true final table compared to the incomplete standings. We have considered a paired-comparison model based on the bivariate Poisson distribution to measure the performance of the teams. Alternative paired-comparison models may have been employed depending on what the target performance measure is. For instance, a paired-comparison ordered probit model could be used to exclude the number of goals scored from the performance measure of the teams. However, we do not expect that our findings will be affected much by such considerations.

References

Bradley, R. A. and Terry, M. E. (1952). Rank analysis of incomplete block designs: I. the method of paired comparisons. *Biometrika*, 39(3/4):324–345.

Buraimo, B., Simmons, R., and Maciaszczyk, M. (2012). Favoritism and referee bias in European soccer: evidence from the Spanish league and the UEFA Champions League. *Contemporary Economic Policy*, 30(3):329–343.

Cattelan, M. (2012). Models for paired comparison data: A review with emphasis on dependent data. *Statistical Science*, 27(3):412–433.

Dixon, M. J. and Coles, S. G. (1997). Modelling association football scores and inefficiencies in the football betting market. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 46(2):265–280.

Goddard, J. (2005). Regression models for forecasting goals and match results in association football. *International Journal of Forecasting*, 21:331–340.

Hvattum, L. M. and Arntzen, H. (2010). Using Elo ratings for match result prediction in association football. *International Journal of Forecasting*, 26:460–470.

Johnson, N. L., Kotz, S., and Balakrishnan, N. (1997). *Discrete multivariate distributions*, volume 165. Wiley New York.

Karlis, D. and Ntzoufras, I. (2003). Analysis of sports data by using bivariate Poisson models. *Journal of the Royal Statistical Society: Series D*, 52(3):381–393.

Karlis, D. and Ntzoufras, I. (2009). Bayesian modelling of football outcomes: using the Skellam's distribution for the goal difference. *IMA Journal of Management Mathematics*, 20:133–145.

Kendall, M. (1938). A new measure of rank correlation. *Biometrika*, 30(1/2):81–89.

Koopman, S. J. and Lit, R. (2015). A dynamic bivariate poisson model for analysing and forecasting match results in the english premier league. *Journal of the Royal Statistical Society. Series A (Statistics in Society)*, 178(1):167–186.

Koopman, S. J. and Lit, R. (2019). Forecasting football match results in national league competitions using score-driven time series models. *International Journal of Forecasting*, 35(2):797–809.

Lit, R. (2020). Time Series Lab - Sports Statistics Edition. https://timeserieslab.com.

Maher, M. J. (1982). Modelling association football scores. *Statistica Neerlandica*, 36(3):109–118.

Pollard, R. (2008). Home advantage in football: A current review of an unsolved puzzle. *The open sports sciences journal*, 1(1):12–14.

Rue, H. and Salvesen, O. (2000). Prediction and retrospective analysis of soccer matches in a league. *The Statistician*, 49(3):399–418.

Some final comments

- All computations and analyses are done by the software package *Time Series Lab - Sports Statistics Edition* of Lit (2020): which is freely available and can be downloaded from `https://timeserieslab.com/`

- The discussions on the premature endings of the 2019/2020 competitions in France and the Netherlands have been widely reported in the media and on social media. Although our study is not taking any position in these discussions, it is interesting that comments in the French newspaper *L'Équipe* made by the president of Lille, Gérard Lopez, can be regarded as relevant to the results presented in Table 2:
`https://www.lequipe.fr/Football/Actualites/Ligue-1-nous-avons-verifie-les-affirmations-de-gerard-lopez-le-president-de-lille/1130690`
`https://www.getfootballnewsfrance.com/2020/gerard-lopez-calls-for-end-of-2019-20-season-to-be-simulated-for-the-sake-of-fairness/`

Online Appendix

This Online Appendix describes the necessary steps to replicate Tables 2 and 3 with the use of the software package *Time Series Lab - Sports Statistics Edition*, hereafter *TSL - SE*. We show how to calculate the model-based prediction of the final standings for the French competition in some simple steps. The model-based construction of the final standings for the Dutch competition is carried out in an analogous manner.

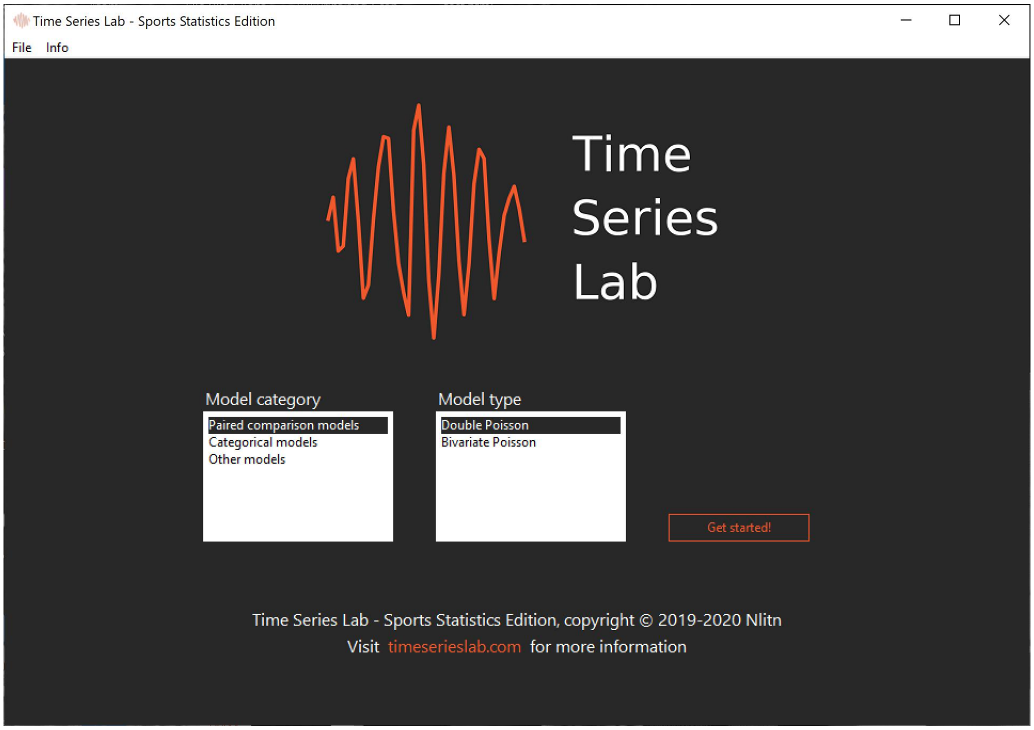
A Installing and starting

The *TSL - SE* software can be downloaded for free from `https://timeserieslab.com`. After installing, *TSL - SE* can be started by double-clicking the icon on the desktop or by clicking the Windows *Start* button and selecting *TSL - SE* from the list of installed programs. The frontpage of *TSL - SE*, which is visible right after the program starts, is shown in Figure 1.

B Loading data

After pressing the *Get started* button on the frontpage you will be taken to the *Load data* step in *TSL - SE*. Click *Load data* and a file selection window opens up. Navigate to the data folder which is located in the same folder where *TSL - SE* is installed. Ctrl-click the files *F11920.csv* and *F11920_remaining.csv* so that both files are highlighted, followed by clicking the *open* button. Alternatively, the data can be downloaded from the Research section of `https://timeserieslab.com`. Once the data is loaded, the screen similar to Figure 2 should appear. An indication that the correct dataset is loaded is given in the upper right corner of the screen. It shows the number of matches per team and the total number of teams in the competition. For the French competition these are 38 and 20, respectively. Since not all scheduled matches were played in the 2019-2020 season, many missing values are part of the dataset.

Figure 1: Frontpage of *Time Series Lab - Sports Statistics Edition*



C Model selection and estimation

Click the *Step 2* button which leads to the *Model setup* page. Select the Bivariate Poisson distribution and tick the boxes in front of *Replace missing values with Expectations* and *Print final table*. A screenshot of the mandatory selections is given in Figure 3

Click the *Step 3* button which leads to the *Estimate* page. Click *Estimate* to start model estimation. After the process of maximizing the likelihood function is completed, output is printed to the Main page of the program. The model-based prediction of the final standings in the French competition is printed on screen as in Figure 4. This printed output matches the results presented in Table 2.

Figure 2: Load data page of *Time Series Lab - Sports Statistics Edition*

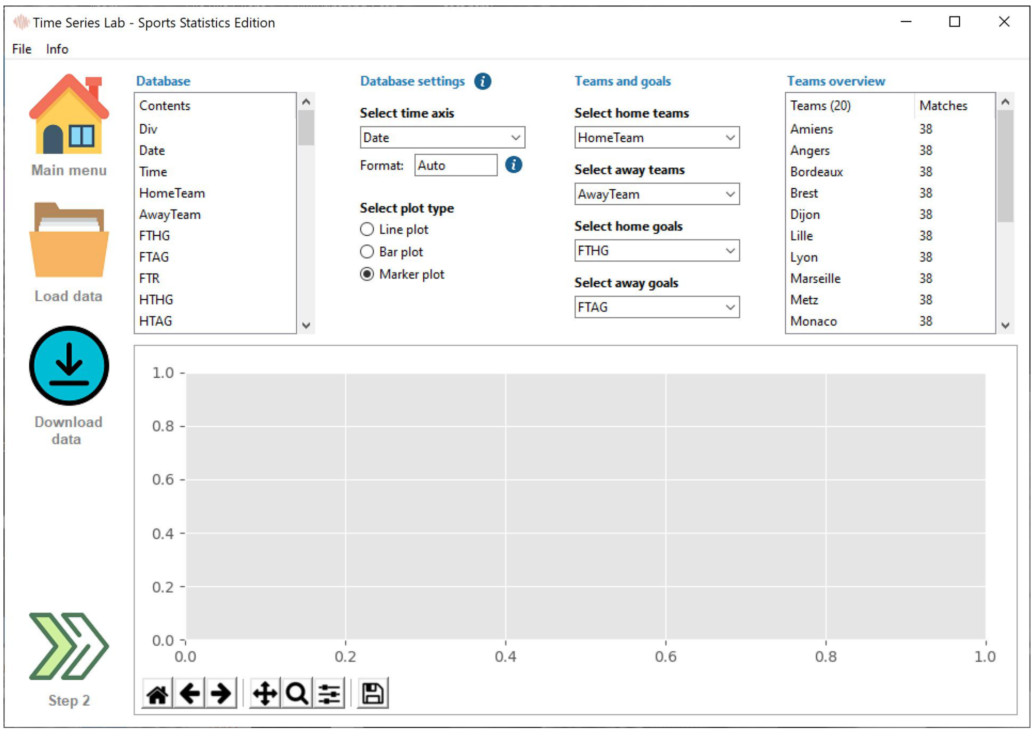


Figure 3: Model setup page of *Time Series Lab - Sports Statistics Edition*

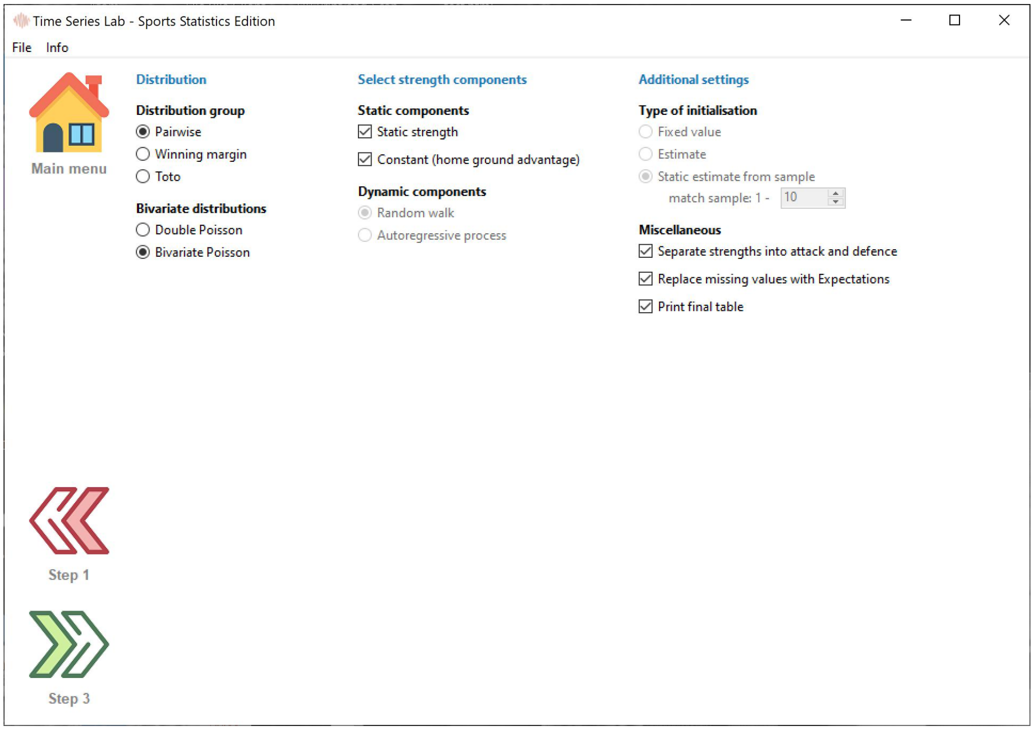
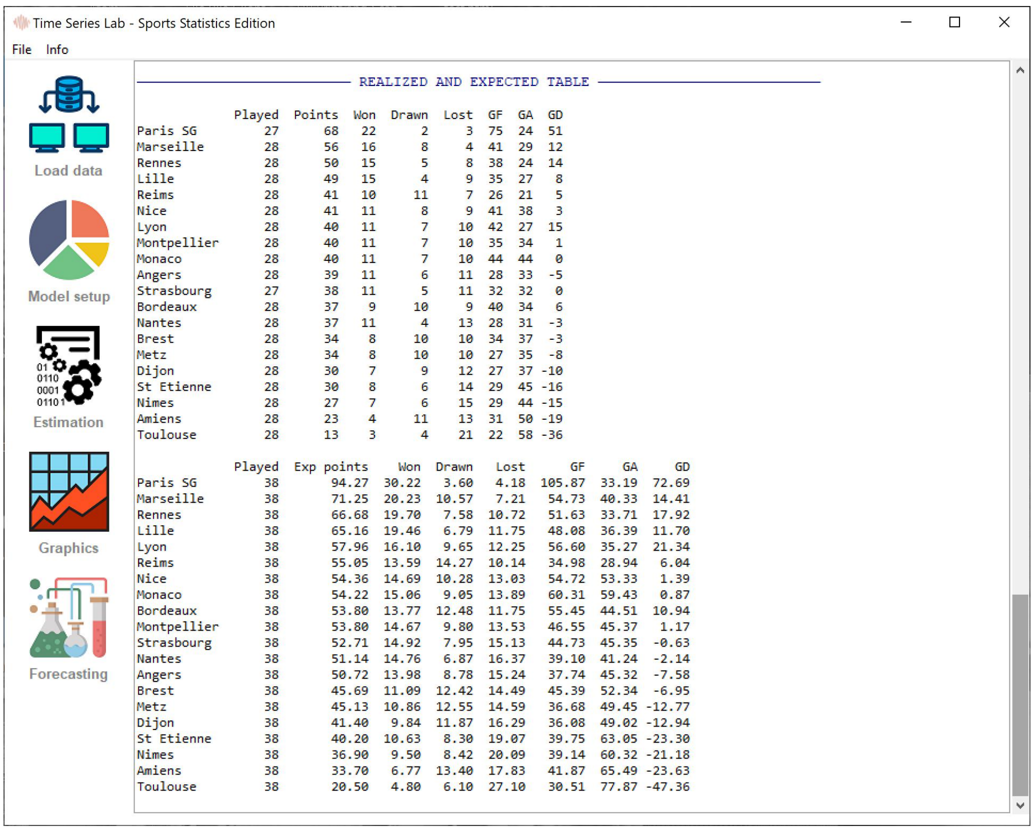


Figure 4: Model-based prediction of the French competition



	Played	Points	Won	Drawn	Lost	GF	GA	GD
Paris SG	27	68	22	2	3	75	24	51
Marseille	28	56	16	8	4	41	29	12
Rennes	28	50	15	5	8	38	24	14
Lille	28	49	15	4	9	35	27	8
Reims	28	41	10	11	7	26	21	5
Nice	28	41	11	8	9	41	38	3
Lyon	28	40	11	7	10	42	27	15
Montpellier	28	40	11	7	10	35	34	1
Monaco	28	40	11	7	10	44	44	0
Angers	28	39	11	6	11	28	33	-5
Strasbourg	27	38	11	5	11	32	32	0
Bordeaux	28	37	9	10	9	40	34	6
Nantes	28	37	11	4	13	28	31	-3
Brest	28	34	8	10	10	34	37	-3
Metz	28	34	8	10	10	27	35	-8
Dijon	28	30	7	9	12	27	37	-10
St Etienne	28	30	8	6	14	29	45	-16
Nimes	28	27	7	6	15	29	44	-15
Amiens	28	23	4	11	13	31	50	-19
Toulouse	28	13	3	4	21	22	58	-36

	Played	Exp points	Won	Drawn	Lost	GF	GA	GD
Paris SG	38	94.27	30.22	3.60	4.18	105.87	33.19	72.69
Marseille	38	71.25	20.23	10.57	7.21	54.73	40.33	14.41
Rennes	38	66.68	19.70	7.58	10.72	51.63	33.71	17.92
Lille	38	65.16	19.46	6.79	11.75	48.08	36.39	11.70
Lyon	38	57.96	16.10	9.65	12.25	56.60	35.27	21.34
Reims	38	55.05	13.59	14.27	10.14	34.98	28.94	6.04
Nice	38	54.36	14.69	10.28	13.03	54.72	53.33	1.39
Monaco	38	54.22	15.06	9.05	13.89	60.31	59.43	0.87
Bordeaux	38	53.80	13.77	12.48	11.75	55.45	44.51	10.94
Montpellier	38	53.80	14.67	9.80	13.53	46.55	45.37	1.17
Strasbourg	38	52.71	14.92	7.95	15.13	44.73	45.35	-0.63
Nantes	38	51.14	14.76	6.87	16.37	39.10	41.24	-2.14
Angers	38	50.72	13.98	8.78	15.24	37.74	45.32	-7.58
Brest	38	45.69	11.09	12.42	14.49	45.39	52.34	-6.95
Metz	38	45.13	10.86	12.55	14.59	36.68	49.45	-12.77
Dijon	38	41.40	9.84	11.87	16.29	36.08	49.02	-12.94
St Etienne	38	40.20	10.63	8.30	19.07	39.75	63.05	-23.30
Nimes	38	36.90	9.50	8.42	20.09	39.14	60.32	-21.18
Amiens	38	33.70	6.77	13.40	17.83	41.87	65.49	-23.63
Toulouse	38	20.50	4.80	6.10	27.10	30.51	77.87	-47.36