

“Stay home” and let the simulation play

Predicting the Outcome Kreisliga A Recklinghausen Season 2019-2020

Working Paper

Submitted to the Faculty of
Economics
at the
University of Duisburg-Essen

from:

Nils Paffen, David Schulze

Reviewer: Prof. Dr. Christoph Hanck

Deadline:

Name:	David Schulze	Nils Paffen
Matriculation Number:	–	3071594
E-Mail:	david.schulze@rgs-econ.de	nils.paffen@stud.uni-due.de
Study Path:	PhD Economics	M.Sc. Economics
Semester:	2 nd	2 nd
Graduation (est.):	–	Winter Term 2021

Contents

List of Figures	II
List of Tables	II
List of Abbreviations	II
1 Literature	1
2 Data	3
3 Predictive Models	3
4 OOSE Test Statistics	7
5 Results	7
6 Conclusion	8

List of Figures

List of Tables

1	Regression ouput of the Poisson Model	6
2	Simulated Final Score Table	8

List of Abbreviations

#Maybe add some sport medical data from an actual research paper how “dangerous” in terms of “enhanced chances of #infection during football games”. I thought I had heard of something like this but will update the section when #I find the paper. # #
Introduction

The Covid-19 epidemic forced sports leagues in Germany to suspend championships that were already in full swing. For example, the Kreisliga Herren Klasse 2 in Recklinghausen finished around 150 games, before the rest were canceled starting from Sunday March 12, 2020, leaving around 90 games left unplayed until the last planned day of the tournament on Sunday May 24, 2020. Given the distinct possibility that there won’t be a chance to make up those games later, a burning question for many players and fans is naturally: What if they would have played those games? We attempt to use data on games already played from the website fussball.de to answer this question, drawing on established forecasting methods from the literature.

1 Literature

Some difficulties with forecasting football matches like that of e.g. the FIFA World Cup, are not relevant to the Kreisliga. The league system in Germany features two rounds per season. Each club plays each other club twice: Once in each round, and once on each club’s home field. This reduces the uncertainty, when compared with the mode of the World Cup. There in the group stage, groups are determined by luck of the draw, a process known as “seeding”. Groups then play a so called round-robin tournament, also known as all-play-all, where every group member plays each other, which corresponds to the mode in which the Kreisliga plays. The World Cup then continues with single-elimination, or a knock-out stage, which introduces another even more random path dependencies that are not needed for forecasting the Kreisliga. This implies that part of the existing literature on forecasting results in the FIFA World Cup concerning the group stage is highly relevant for the task at hand, since the game rules are identical.

A natural starting point for forecasting match or season outcomes in football tournaments is using the FIFA points ranking method that is used canonically to evaluate the strength of a team and updated after each game. For example, a recent study by Correa et al. (2018) uses FIFA points to forecast the results of the 2018 FIFA Men’s World Cup. This approach has however generated criticism and other methods have been proposed and evaluated. The benchmark study by Laseck et al. (2013) compares established and proposed rankings. They find that FIFA rankings used to perform slightly worse than alternative methods, especially two versions of the Elo rating system originally proposed

by Arpad Elo for the United States Chess Federation to rate competitive chess players. We will consider using the FIFA points method or these two alternative candidates.

The first candidate model is published by the anonymous website EloRatings.net (2012), and is especially adapted for the use in ranking football teams. The second one is the FIFA Women’s World Rankings, which has been in use since 2003. It is worth noting that the FIFA Men’s World Rankings have been adapted in 2018 to also be based on a modification of the Elo ranking.

The website FuPa.net (2020) also publishes detailed statistics as well as a so called “Power-Ranking” to evaluate the “Formkurve”, a rough measure of very recent performance. Because it is only reflective of very recent games, it is not useful for long term forecasting. Another raw measure would be to calculate the probability of winning by dividing team’s current points (victories are 3, draws are 2) by the total of their and their opponent’s points, we can call this the “points model”.

All these models just use weighted game results or point scores. Another approach would be to use more data, as in Berbée et al. (2015). While this concerns baseball, the principle is not dependent on the rules of the game, but it is based on the influence of individuals’ and team’s statistics on game outcomes. This is an application of the model proposed by Albert and Bennet (2007). They use the fact that wins and losses appear to be normally distributed in the long run, so chance and skill should both play a role. Team ability is then calculated as the deviation from the average winning probability over time. Very generally, available statistics are then evaluated to find those that matter the most for team performance, measured in runs per game, which would be goals in football. Those candidate factors are then weighted by regressing them on the team performance. The resulting index is then used to calculate a winning probability and this used then to simulate the season outcome by using weighted coin-flips.

While there is not as much individual player data in football on the level of our analysis, the team average is still something that can be calculated. Also, more information on the games, such as location at home or away, could be candidate factors for performance. For example, one team might benefit more from playing at home than it hurts by being away, receiving a net benefit from location, or the other way around. While these data are certainly interesting, we must be aware of the danger of overfitting the model.

Finally the literature contains references to using betting markets as a benchmark, as they perform relatively well as predictors. For the Kreisliga however, we don’t expect betting markets to be deep enough to reach this level of accuracy. Otherwise, they would also be an interesting reference point.

2 Data

For the simulation study we decided to use data from the amateur football league of the Kreisliga A Kreis Recklinghausen class A1 in Westphalia, further called Kreisliga A1, for the 2019/2020 season. 16 clubs will play against each other on a total of 30 match days in one home and one away game each. Due to the Covid-19 pandemic, the association has decided to cancel all matches from March 15, 2020. After all, 20 matchdays have already been played until this point in time which corresponds to a database of 158 matches. As the first half of the season had already been completed, each team had already played at least once against each team in the table. The extraction of real data from websites using scraping scripts can be complicated, as website operators have an interest in protecting their data from such automated queries. “Fussball.de” is a website of the DFB (German Football Association) which acts as a collection point for match results and news, especially in the amateur sector. The match results of the website itself cannot be directly read out. They are masked, so they are made unreadable when viewing the HTML file and are only evaluated afterwards using Javascript and transferred to the CSS of the site. The site also offers a match report, which graphically represents a temporal course of the match. This is broken down in the HTML code, in contrast to the match results, unmasked, and shows the course of the match in text form. With the help of regular expression operations, the game result can be reconstructed. The data record was then divided into completed and unplayed games. The latter amount to 89 in this season, which were simulated with the methods in the following chapters.

For further analysis we decided to scrape the data of season 16/17, 17/18 and 18/19 aswell to perform out-of-sample error (OOSE) test statistics. The latter will indicate the performance of the different methods.

Give a short overview of the actual standing (ranking table after matchday 20)

3 Predictive Models

To predict the outcome of the cancelled games, we calculate the candidate rankings and use them to simulate the end of the 2019/2020 season, usually by way of calculating a winning probability for each missing game. Specifically we calculate

- the points model,
- the EloRankings.net model,
- a poisson model

The first model is just a simple baseline model that calculates the probability of a team A winning a game against team B using the formula

$$P(A \text{ wins}) = \frac{\text{tablepoints}_A}{\text{tablepoints}_A + \text{tablepoints}_B}, \quad (1)$$

where *table points* corresponds to the number of games won at the current state of the season valued at three points plus the number of draws valued at one point. This value also governs the ranking and ultimate placement of the teams in the league. Two questions arise from this approach. Firstly, should the table points be updated after each simulated game? We argue no, because this would not include new information about the relative strength of the teams and just increases the variance of the result. Secondly, we expect an average over many simulated runs to converge to the initial table when the season was interrupted. This would then defeat the purpose of running a simulation in the first place, because it does not add any new information, and we could have just used the table as it were. Using this way of simulation is however preferable to an unweighted coin toss, because that would unfairly favor below average teams.

Our second model is based on the rating algorithm from elratings.net. The anonymous site operator formulates the rating, representative of the strength of a team, as follows:

$$R_n = R_0 + K \times (W - W_e). \quad (2)$$

Here, R_n is defined as the new rating as an update of R_0 , which is the old rating. The weighting factor for each match is defined by the type of tournament in which the match takes place and also controls for friendly matches, which is given to the lowest weight of 20. While matches in world championships and other major international tournaments are given weights between 40 and 60, the rest falls into the category “all other tournaments” which are given a weighting factor of 30. Following this example, we also set K to 30 for matches already played in the Kreisliga A. The weighting factor K is adjusted again based on the goal difference of the result. Thus, K is increased by $\frac{K}{2}$ if the match was won with two goals, by $\frac{3}{4} \times K$ if the match was won with three goals and by $\frac{3}{4} + \frac{(N-3)}{8} \times K$, where N defines the goal difference of the match if the match was won with four or more goals. W is the result of the match. 0 for a loss, 0.5 for a draw and 1 for a win. W_e is the probability of winning defined by the following formula:

$$W_e = 1/(10^{(-dr/400)} + 1) \quad (3)$$

,

where dr is defined as the rating difference and the home team receives a bonus of 100 points. This bonus is considered to be a psychological advantage resulting from the fact that the game is played in the home stadium(see, e.g., Pollard (2008)).

Our third model uses the poisson distribution to simulate the match result with the probability of a goal in every minute of a match. The probability matrix where the game result is drawn from is a $n \times n$ matrix where each cell indicates the probability of that specific match result. While the rows indicates the goals of the home team, the column indicates the goals of the away team. For example the cell of the first row and in the first column indicates the likelihood that the both teams score 0 goals. The poisson probability function of our model can be expressed as:

$$P(x) = \frac{e^{-\lambda} \lambda^x}{x!}, \lambda > 0 \quad (4)$$

where the λ represents the average number of goals. First, we estimate the following model from the matches already played:

$$goals = home + team + opponent. \quad (5)$$

Where $goals$ represents the number of goals scored by a team in a game, $home$ presents a dummy variable that is 1 for the home team and 0 for the away team, it serves to illustrate the home team bonus, $team$ representing the home team, and $opponent$ representing the opponent team.

% Table created by stargazer v.5.2.2 by Marek Hlavac, Harvard University. E-mail: hlavac at fas.harvard.edu % Date and time: Fr, Aug 21, 2020 - 12:06:12 % Requires LaTeX packages: dcolumn

In summary, the coefficients of the model show that the club “Altendorf-Ulfkotte”, both as home and away club, has a strong negative and a strong positive influence???, both highly significant, on goals, i.e. the number of goals. Since the club is in the last place in the current table, as mentioned in the Data section, we expected that it would be easier to score goals if they played against the team on the last place of the seasons

Table 1: Regression output of the Poisson Model

	control	goals team	opponent
	(1)	(2)	(3)
Constant	0.752*** (0.228)		
home	0.241*** (0.076)		
Adler Weseke II		-1.047*** (0.249)	0.595*** (0.218)
BVH Dorsten		-0.289 (0.199)	0.051 (0.250)
FC RW Dorsten		-0.877*** (0.232)	0.178 (0.237)
Fenerbahce I. Marl		-0.564*** (0.206)	0.109 (0.244)
SC Marl-Hamm		-0.145 (0.192)	0.507** (0.226)
SC Reken II		-0.405** (0.206)	0.697*** (0.220)
SV Altendorf-Ulfkotte		-1.252*** (0.277)	1.089*** (0.205)
SV Lembeck		-0.216 (0.196)	0.356 (0.230)
SV Schermbeck II		-0.167 (0.186)	-0.267 (0.272)
TSV Raesfeld		0.021 (0.179)	-0.085 (0.258)
TuS 05 Sinsen II		-0.902*** (0.241)	0.581*** (0.219)
TuS Gahlen		-0.266 (0.191)	-0.812*** (0.315)
TuS Velen		-0.409** (0.202)	0.280 (0.233)
VfL Ramsdorf		0.072 (0.177)	-0.435 (0.283)
Westfalia Gemen II		-0.559*** (0.210)	0.591*** (0.220)

Notes: ***Significant at the 1 percent level.

**Significant at the 5 percent level.

*Significant at the 10 percent level.

soccer league table rather than the team on the first place. On the other hand it will be harder for this last placed team to score goals even if they are the home team.

Following Correa et al. (2018), we then run the simulation by drawing the results of each game from a binomial distribution. For each game and team, the probability of winning is dividing the ranking points awarded each team by their and their competitors sum of points.

Running this simulation repeatedly should indicate the distribution and expected average of outcomes. Correa et al. (2018) execute 200,000 runs, but because of the relatively low complexity of the Kreisliga's format compared to the World Cup, especially because there are no elimination rounds, we expect to need less repetitions.

A few alternatives have been developed for forecasting football games. The potential of using independent Poisson distributions to match the empirical distribution of goals scored by a team has been improved on by introducing correlation between the teams playing against one another in a bivariate Poisson distribution Karlis and Ntzoufras (2003).

While the independent Poisson distributions already allowed for a better fit and to model the outcome of draws, Boshnakov et al. (2017) used a Weibull count model to improve even on the bivariate Poisson model, allowing them even to outperform betting market in selected bets.

4 OOSE Test Statistics

Making predictions of events that might never happen can fairly be criticized by a simple question. How do you know that your results reflect reality as good as possible? Following George E. P. Box who is known for his quote "All models are wrong" which is often extended by "but some are useful" we want to show that our models cover the latter. The out-of-sample error test statistic is one way to achieve this. One simply divides a dataset into a small test data set and a larger training data set. For the seasons 16/17, 17/18 and 18/19 we decided to split the dataset at the same point where the COVID-19 pandemic forced the

5 Results

For the simulation study using the elo rating, as explained in the predictive models chapter, we used the average of all matches played in the current season resulting in a tie for the probability of a draw. Half of the percentage points are deducted from the

home team's winning probability and half from the away team's winning probability. Then we draw from these three probabilities the game result, team home wins, team away wins or draw. We repeat this procedure for all games and evaluate the results with 3 points for the winning team, 1 point for both teams in case of a draw and 0 points for the losing team.

In the poissonmodel we calculate for each match the goal probabilities of both teams as a probability matrix based on the model estimation as described in predictive models part.

All simulations were repeated until the rate of change of the point average was 1% or less. Aggregation to this point occurred after about 2580 for the elo model and after about 1980 for the poisson model.

Table 2: Simulated Final Score Table

rank	Poisson Distribution Model		Elo Rating Model		Points Model	
	club_name_poisson	score_poisson	club_name_elo	score_elo	club_name_points	score_points
1	VfL Ramsdorf	63.46088	VfL Ramsdorf	64.34676	VfL Ramsdorf	72.19792
2	TuS Gahlen	57.54770	TuS Gahlen	58.74429	TuS Gahlen	65.04059
3	SV Schermbeck II	56.68854	SV Schermbeck II	57.74274	SV Schermbeck II	63.51325
4	Fenerbahce I. Marl	52.45432	Fenerbahce I. Marl	53.52383	Fenerbahce I. Marl	58.06307
5	1. SC BW Wulfen	49.74609	1. SC BW Wulfen	50.88454	1. SC BW Wulfen	57.22107
6	TSV Raesfeld	49.58253	TSV Raesfeld	50.64084	TSV Raesfeld	55.75243
7	TuS Velen	40.56790	TuS Velen	41.79078	TuS Velen	45.37873
8	SC Marl-Hamm	39.67895	SC Marl-Hamm	40.90701	SV Lembeck	44.14961
9	BVH Dorsten	38.56991	BVH Dorsten	39.89500	SC Marl-Hamm	43.87890
10	SV Lembeck	38.49319	SV Lembeck	39.56838	BVH Dorsten	42.64140
11	FC RW Dorsten	34.61080	FC RW Dorsten	35.82449	FC RW Dorsten	37.59242
12	Westfalia Gemen II	31.46037	Westfalia Gemen II	32.40101	Westfalia Gemen II	32.82422
13	SC Reken II	27.79455	SC Reken II	29.11275	SC Reken II	27.60483
14	TuS 05 Sinsen II	22.68854	TuS 05 Sinsen II	23.86207	TuS 05 Sinsen II	21.27809
15	Adler Weseke II	21.69107	Adler Weseke II	22.59512	Adler Weseke II	18.88125
16	SV Altendorf-Ulfkotte	18.52751	SV Altendorf-Ulfkotte	19.50446	SV Altendorf-Ulfkotte	13.98222

6 Conclusion

The decision to quit all games later than 08th of march because of the pandemic was not revised while the infection rates relaxed during may and june in Germany. Combined with the unforeseeable future of the COVID-19 situation we see a more fair and balanced decision making process by integrating statistical learning techniques, such as those, shown in this paper.

TO DO:

-clean up models section: use more original sources, Maher 1982? -calculate draw in poisson from diagonal? -add caveats: what other factors play a role, what alternative models are there? use sources -goal distribution?

References

- Albert, J., & Bennett, J. (2007). *Curve ball: Baseball, statistics, and the role of chance in the game*. Springer Science & Business Media.
- Boshnakov, G., Kharrat, T., & McHale, I. G. (2017). A bivariate weibull count model for forecasting association football scores. *International Journal of Forecasting*, 33(2), 458–466. <https://doi.org/https://doi.org/10.1016/j.ijforecast.2016.11.006>
- Correa, M., Barrera-Causil, C., & Marmolejo-Ramos, F. (2018). The next winner of the 2018 fifa world cup will be...: An illustration of the use of statistical simulation to make a prediction in a complex tournament. *Chilean Journal of Statistics (ChJS)*, 9(1).
- Fupa.net [Accessed: 2020-04-16]. (2020).
- Karlis, D., & Ntzoufras, I. (2003). Analysis of sports data by using bivariate poisson models. *Journal of the Royal Statistical Society: Series D (The Statistician)*, 52(3), 381–393. <https://doi.org/10.1111/1467-9884.00366>
- Lasek, J., Szlavik, Z., & Bhulai, S. (2013). The predictive power of ranking systems in association football. *International Journal of Applied Pattern Recognition*, 1(1), 27–46.
- Lobeck, M., Schulze, D., & Berbee, P. (2015). Can he hit? measuring and prognosis of offensive performance by baseball teams and individual teammembers. *Seminar Paper, B.Sc. VWL HU Berlin*.
- Pollard, R. (2008). Home advantage in football: A current review of an unsolved puzzle. *The Open Sports Sciences Journal*, 1. <https://doi.org/10.2174/1875399X00801010012>
- The world football elo rating system [Accessed: 2012-03-03]. (2012).