# Predicting the Premier League

Ben Folland, Callum Knight, Alex Maksymczuk, Edmund Fisher

Exeter Mathematics School

Author Note

ATASS Sports, Dr Tim Paulden

## Abstract

This is an investigation of how effective simple mathematics can be in predicting the results in one of the most competitive competitions in the world: the Premier League. We did this because it offers the opportunity for casual football fans to compete with the bookmakers, and others, who employ vast amounts of data, processed in a variety of ways to set odds. Inspired by the work of David Spiegelhalter, we sought to improve his results by refining his method, which is based on the calculation of expected goals, using goals scored and conceded at home and away. We applied this to the results in this season's Premier League from 3$^{rd}$ April – 23$^{rd}$ April. We improved Spiegelhalter's model slightly by altering the amount of emphasis put on home and away performances, but still the model was not fantastic, with a generally low rate of correct score lines predicted. This is because football is not generally a predictable sport, with lots of external factors affecting every game that are extremely difficult to model mathematically. The dream of a simple mechanism to beat the bookies remains elusive.

## Introduction

Throughout human history attempts have been made to predict results for seemingly unpredictable events. Gambling has a long history, in many different cultures around the world, and people have used logical reasoning and mathematical formulae to try and improve their chances of winning. In ancient Rome for example predictive games were very popular and our archaeological record has produced evidence of forms of dice, other games and even gambling tables. Chariot racing was also accompanied by betting on the results. (Reith, 1999) By the third century AD the Circus Maximus had become a centre for elaborate gambling with over a million spectators in one stadium watching city based teams compete against each other (Ingrid de Haas, blog, no date). Over time, techniques for predicting results of sporting events have been refined. Some of the most complex are employed by bookmakers to help them determine odds for bets. Recently, in this new age of computers, complex programs and algorithms have been developed to try and apply a more complex mathematical approach to predicting results. We were asked to create a program on R Studio that took data from a Premier League season and found the most likely score line for any matchup, using Spiegelhalter prediction method. We were given all the Premier League data, and the Excel spreadsheet that David Spiegelhalter used to implement his prediction process.

## Methodology

We sought to investigate the possibility of accurately predicting Premier League results by applying David Spiegelhalter's logic based formula to a substantial number of results over several years. David Spiegelhalter, Winton Professor of the Public Understanding of Risk in the Statistical Laboratory at the University of Cambridge, developed his model in 2009 to complete a challenge set by BBC Radio 4's More or Less, to predict the results for the final 10 matches of the 2008/09 season using a method that could be explained to non-specialist listeners. He later used the same method to predict results during the 2009/10 season, which was dubbed "the most unpredictable ever" (Spiegelhalter, 2010). The appeal of this method, which informs our choice to conduct this experiment, is that it does not rely on the enormously complex mathematical models employed for example by bookmakers, and offers the potential to provide a simple, understandable, and accurate method for predicting results. As he explained: "Complex statistical models are used extensively in the sports betting industry, but we wanted a fairly straightforward model that could be explained using familiar concepts derived from the raw data" (Spiegelhalter, 2009). His method focuses on the notion of 'attack strength' which is a measure calculated by comparing numbers of goals scored by a particular team against the average scored in the premiership as a whole. The measure of 'attack strength' is compared with a second measure, that Spiegelhalter labels 'defence weakness', which is calculated by comparing numbers of goals conceded by a particular team in relation to the average conceded by all the teams in the premiership as a whole. Finally, these measures are modified by two additional pieces of information. Each team's attack strength is split into separate scores for their home and away performances. Having produced these calculations Spiegelhalter can start the predicting process. First, he produces a value for the number of expected goals scored by the home team by combining the home attack strength, the away defence weakness and a particular weighting reflecting home/away advantage. This weighting is changeable and based on human judgement on the qualitative knowledge of recent matches. However, these values for expected goals will nearly always be decimals, and no team can score, for example, 0.4 of a goal. Therefore, these expected goals are distributed onto a Poisson distribution, to find the probability for a range of score lines, the most probable of these taken as the final prediction. The Poisson distribution is a probability distribution that shows the probability of a few differing results occurring. For the case of predicting results, by inputting the expected goals into the Poisson distribution, you can find the probability of each team scoring a certain number of goals, and then by combining these values in a table to find the most probable result.

Spiegelhalter's method differs from other approaches to the science of predicting sporting events, which is a widely researched area of mathematics. For example, Dimitris Karlis and Ioannis Ntzoufras developed a different and more complex method for predicting results. First, their method placed emphasis on the goals difference within each game, instead of the number of goals scored which is favoured by Spiegelhalter. They placed particular value on the margin of victory as a means of understanding the strength of any team and use this variable to model a team's likelihood of victory in forthcoming matches. Second, they apply Bayesian model averaging techniques to their model to improve its predictive power; this iron out uncertainties in the input data (e.g. skews caused by weather conditions, or the fitness and injuries within the team) and uncertainties in the output. Another group of mathematicians, Leonardo Egidi, Francesco Pauli, and Nicola Torelli, developed a different method, also involving Bayesian modelling, but additionally incorporated the Poisson distribution (also used by Spiegelhalter), and using data from the previous 9 years to improve accuracy via a larger sample size.

Despite the value and usefulness of these interesting approaches, the beauty of Spiegelhalter's model is its simplicity. If an accurate but simple model for predicting results could be established, then the fun of predicting football games could be enjoyed by people who do not have access to the expertise, computer hardware and mathematical knowledge required to perform these complex, difficult and data-heavy calculations. Our project, therefore, is conceived as an attempt to analyse, test, and modify Spiegelhalter's method, in pursuit of a simple but relatively plausible method for predicting football results – one that would be enough to risk pitting against the expert bookmakers; one that provides a non-mathematically expert punter with a decent chance of profitable return.

The core of our method was the same as David Spiegelhalter's and used the same model. However, we altered the weighting, in an attempt to increase the accuracy of the predictions. Spiegelhalter's model had only a success rate of around 52% for the correct outcome (which team would win/draw), and roughly 8% for exactly the right score line. So, we decided to alter the weighting to see if we could find a more accurate one, and find an increase in the percentage of correct results and score lines. One overall aim of the project was to refine and improve Spiegelhalter's model, maintaining the simplicity, but improving the accuracy via modifying the weighting. We did this by trying out a series of different weightings from 0 to 1. We chose the weighting produced the highest number of correct results and score lines and picked this as our best weighting for this season. This ended up being the weighting of 0.5, which means that no extra emphasis is put on home or away attack strength or defence weakness. An explanation for this could be that due to COVID-19 no fans are allowed inside the stadiums to support their teams and cheer them on, something which many footballers and pundits say give you the extra few percent you need to create one extra chance, and even maybe a goal, and plays a huge role in home advantage. This method of testing weighting providing a mechanism for understanding the significance of home advantage in non-COVID situations.

Having chosen which weighting method to apply, the second stage of our project was to apply our chosen weighting to the method in the following way. Our chosen weighting was applied to the formula to create a new value for attack strength and defence weakness. This involved multiplying the average goals scored at home by 1- the weight and the average goals scored away by just the weight, before dividing it all by the opposite (where the 1- the weight is multiplied by the away goals scored and just the weight by the goals scored at home, so that when the weighting is toward the home team, more emphasis is put on a team's home attack than its away attack. The same thing is done with the goals conceded to find the weighted defence weakness. The weighted attack strength for the home team is multiplied by the weighted defence weakness for the away team, and then multiplied by the average goals scored by all teams in the premier league at home. This produces the value for expected goals. As described above, this is then inputted into the Poisson distribution and the probabilities multiplied together to find all the percentage probabilities of every outcome, the highest percentage taken as the prediction.

| | | | | **Arsenal** | | | |
|---|---|---|---|---|---|---|---|
| | | 0 | 1 | 2 | 3 | 4 | 5+ |
| | 0 | 6 | 13 | 15 | 11 | 6 | 4 |
| | 1 | 4 | 8 | 9 | 6 | 4 | 2 |
| **Fulham** | 2 | 1 | 3 | 3 | 2 | 1 | 1 |
| | 3 | 0 | 0.5 | 1 | 0 | 0 | 0 |
| | 4 | 0 | 0 | 0 | 0 | 0 | 0 |
| | 5+ | 0 | 0 | 0 | 0 | 0 | 0 |

So, for example in case of a match between Fulham and Arsenal, as the above chart shows, the most likely score line is 2-0 to Arsenal (the home team) with a 15% chance of this occurring.

Throughout this process we used R Studio to process the data using the algorithm that we created. R Studio is a frequently used programming language for statistical analysis and data processing that that incorporates the key functions we wanted to apply. It was devised to format abstract statistical data and produce results including predictions. We transferred the raw data from David Spiegelhalter's excel spreadsheet and coded his method for finding the most likely score line for every game. We ran data through RStudio to test weightings, as described above, and edited the code to incorporate our chosen weighting. New data was processed through this new algorithm, including the application of Poisson distribution to provide realistic and viable predicted results. This gave us a predicted result with probabilities for every game.

Our model was then complete and only required the number of home/away goals scored/conceded by every team in the league up to that point, and the number of games played to come up with a prediction. It was, as designed, simple, logical, and straightforward.

# Results

We were successful in one of our aims. Our refinement of David Spiegelhalter's model did improve its predictive abilities. We increased the accuracy rate of our model by simply changing the weighting from David Spiegelhalter's 0.33, to 0.5. This effect was only significant in one measure of accuracy. Both our model and David Spiegelhalter's produced similar rates regarding the final outcome of a match (i.e. who won, or if it was a draw). Both models predict the right winner around 52% of the time. Another measure of accuracy – the precise score line was much more regularly correct in our model than David Spiegelhalter's. Indeed, the percentage of correct exact score lines increased by 87% (from Spiegelhalter's 8% to our 15%). However, we cannot conclude that our model is anything like an effective or reliable predictor of football results/scores. The general unpredictability of football as a game is the key limitation of these models. Football does not operate on a mathematical basis. It is subject to huge amounts of individual and circumstantial variation which defy simple mathematical modelling. This particularly clear when it comes to predicting precise score lines or individual matches.

*Our Weighting:*

We tested lots of different weights on a number of matches. An example of the calculation as applied to the matches in Gameweek 30 ($3^{rd}$ April – $5^{th}$ April), is below. This chart shows us testing the following weightings: 0, 0.1, 0.25, 0.33, 0.5, 0.75, and 1.

| Weight | 0 | 0.1 | 0.25 | 0.33 | 0.5 | 0.75 | 1 |
|---|---|---|---|---|---|---|---|
| Correct score line predicted | 2 | 2 | 1 | 1 | 1 | 0 | 0 |
| Correct outcome predicted | 5 | 7 | 6 | 6 | 7 | 6 | 4 |

Although from these results we might have concluded that 0.1 would be the best weighting, it was not the best across the board. On the basis of doing this exercise with many games from this season, we deduced that 0.5 would be the most appropriate weighting; it was the one that most consistently provides the most accurate results compared to the other weightings.

Another interesting observation we made is that our model is much better at predicting correct score lines when it was a lower scoring match. It was especially poor at predicting high scoring matches, such as the same between West Ham and Leicester in which the result was 3 – 2 Leicester, but which our model predicted to be 1-1. When we ran David Spiegelhalter's model over 110 random games, it also shared this property, only predicting one match out of the 110 to have 5 or more goals. Our model is therefore skewed towards predicting low scores. This is because on a Poisson distribution, there is generally a low probability of scoring high numbers of goals, which is also generally true for most football matches, the high scoring ones being outliers.

*Some examples of our predictions:*

For Gameweek 30 we predicted 10 games, 7 correct score lines, and 1 exact score.

However, there was one huge outlier in this prediction, as the actual result for the Chelsea West Brom game (which we predicted to be 2-0 to Chelsea) was Chelsea 2-5 West Brom. According to our model, there was approximately a 0% chance of West Brom scoring 4 or more goals, yet it happened.

*1 A chart showing the different probabilities (%) for each score line.*

| | | Chelsea | | | | | |
|---|---|---|---|---|---|---|---|
| | | *0* | *1* | *2* | *3* | *4* | *5+* |
| | *0* | 6 | 14 | 17 | 13 | 8 | 6 |
| | *1* | 3 | 6 | 8 | 6 | 3 | 3 |
| **West Brom** | *2* | 1 | 1 | 2 | 1 | 1 | 1 |
| | *3* | 0 | 0.2 | 0 | 0 | 0 | 0 |
| | *4* | 0 | 0 | 0 | 0 | 0 | 0 |
| | *5+* | 0 | 0 | 0 | 0 | 0 | 0 |

This is an example of the unpredictability of football and the ways in which a wide host of factors contribute to any result. Although Chelsea took the lead, they were reduced to 10 men when one of their defenders was sent off in the 29$^{th}$ minute, giving West Brom 61 minutes to play 11 vs 10, presenting them with a significant advantage. This advantage was clearly at play in the unusual result and is precisely the type of factor that our model does not even try to accommodate. Even though it is an event that makes the result an outlier, these types of events are in fact relatively common in football.

In addition to being ill-equipped to handle the unpredictability of football as a game, there are other assumptions within the model that can skew the results, or aspects of football as a game that it does not hand. The model does not take into account the different abilities of both teams in a match. Teams are assessed according to their form in recent matches which is not assessed in relation to the opposition they faced. It is therefore only applicable to league matches in which all the teams are playing each other and doing so in a league that is controlled for quality. As a result, the ability of all teams must be relatively equally and roughly comparable. Our model would not be applicable at all for example in FA Cup games where teams from different leagues meet each other. Here the allocated values for attack strength and defence weakness would be skewed by their previous, different, game experiences. If we tried to apply the model to such an event the results would be highly likely to be anomalous. Many poor teams would be predicted to beat much better teams. A match between a high-ranking lower league team with a history of scoring many goals (against relatively weak opponents) would be predicted to win, and score heavily, against a team that were struggling but playing in a higher league. This is not likely to happen. Furthermore, the model cannot consider complex variables, such as player injuries, red cards, a change of manager, or newly signed players, despite the fact that anecdotal evidence would suggest that these have a significant impact on the outcomes of games.

*Conclusion*

In conclusion, our investigation has demonstrated the challenges involved in predicting highly variable events such as football matches. Despite the opportunities provided by the ability of computer programs to handle large amounts of data, the difficulties remain. Our method, a refined version of David Spiegelhalter's prediction model, although it improved on his performance slightly, remained a poor guide to football results and scores. The model did not consider a number of key variables that affected matches, and overall, the range of events and processes that dramatically shape games and determine their result, make football an unpredictable game.

## Bibliography

Ingrid de Haas, *In Rome, all was fair in games and races,* The Ultimate history project, https://ultimatehistoryproject.com/in-rome-all-was-fair-in-games-and-races.html

Gerda Reith, *The Age of Chance: Gambling in Western Culture,* Routledge, 1999.

Spiegelhalter, D. and Ng, Y.-L. (2009), One match to go!. *Significance*, 6: 151-153. https://doi.org/10.1111/j.1740-9713.2009.00387.x

David Spiegelhalter, 'Predicting the Premiership', BBC News, 8 May 2010,

*http://news.bbc.co.uk/today/hi/today/newsid_8668000/8668153.stm*

## Appendices

Our whole project code and results,

https://github.com/polygonben/Predicting-the-Premier-Leauge

.