English Premier League Monte Carlo Analysis

Megan Dunnahoo, Jasmine DeMeyer, Macey Dodd

12/16/2021

Article

We chose the article "Using Monte Carlo Simulation to Calculate Match Importance:

The Case of English Premier League" by Jiri Lahvicka. This article describes the process of using Monte Carlo simulations to predict the outcome of a match given the results of previous matches. It specifically predicts the result of the Manchester City versus the Manchester United game in 2012. It then goes further and uses Monte Carlo simulations to predict the final ranking of the teams in the English Premier League at the end of a season.

Background Information

The English Premier League is regarded as the most popular sports league in the world due to its massive audience views and impressive revenue. There are 20 teams in the English Premier League. Manchester United is considered to be the most popular football club with the Liverpool club in second. In football, a game can result in a tie as well as a win or a loss. Three points are awarded for a win, one for a draw and zero for a loss. At the end of each season, the lowest ranking three teams will be "relegated" or demoted to the lower football league, the English Football League (EFL). The highest three ranking clubs in the English Football League will be "promoted" into the Premier League.

Data and Code Setup

We got our data from football-data.co.uk. We wanted to use the specific variables FTR, FTAG, and FTHG, along with the identifier variables of Date, Away Team, and Home Team. There were no NA values in any of our selected variables. We chose to use four seasons ranging from the years 2011 to 2015. Each team played 19 away games and 19 home games per season.

We created functions to get the points scored and the outcome of the team for each team over the four years, specifying away or home games for each team.

Monte Carlo Estimation

We estimated the lambda home and lambda away values, which are the expected goals scored by the home and away teams respectively, using Monte Carlo. These lambda values are assumed to be independent Poisson distributed variables and are calculated using the last 19 matches for each team. The article ran 10,000,000 simulations, but due to our low computational power, we chose to run 100,000 simulations. The general purpose of Monte Carlo is to model the probability of different outcomes and reduce uncertainty. It is very useful for modeling probabilities that come from processes in which random variables intervene with each other, making them difficult to predict. The purpose of the paper's Monte Carlo simulations is to devise a new way to calculate match importance which refers to the relationship between match results and the final outcome for a specific season.

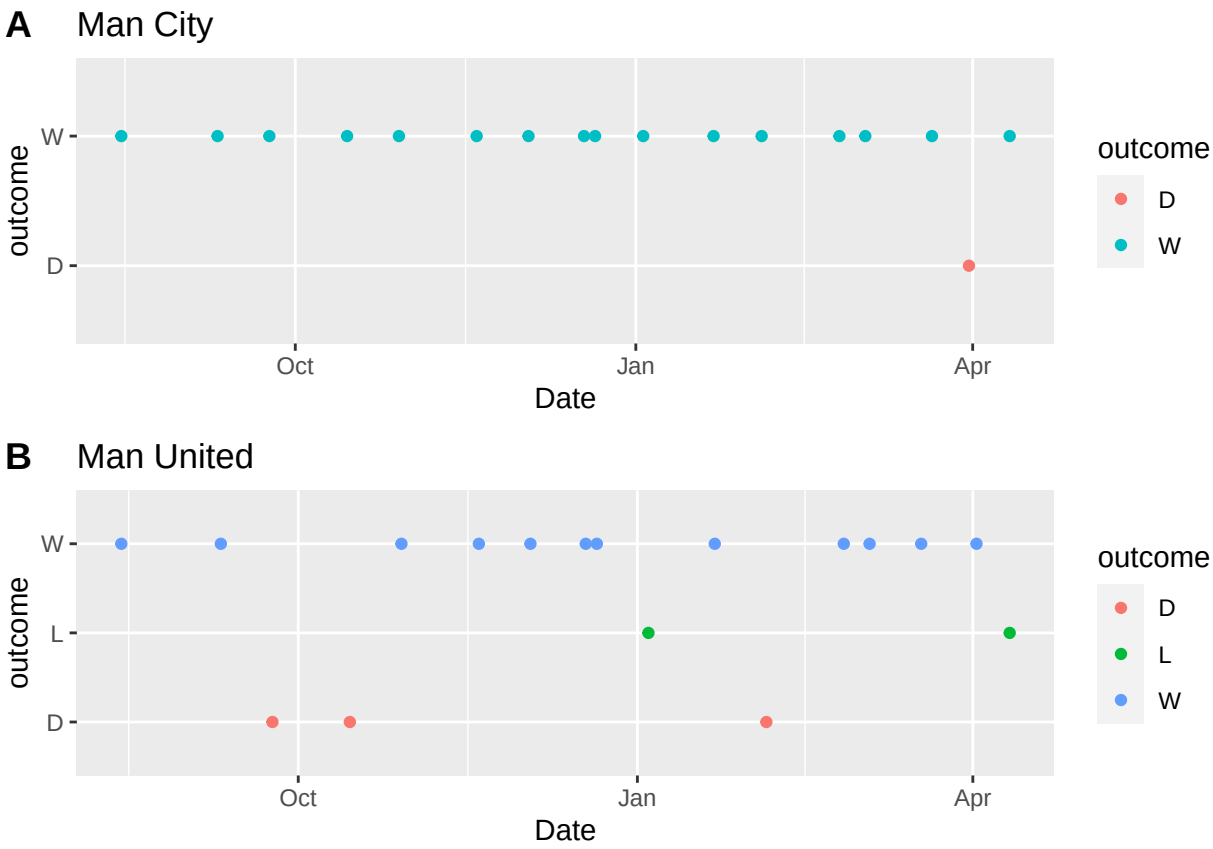
$$\lambda_{home} = \frac{\text{Average goals scored by home team} + \text{Average goals conceded by away team}}{2}$$
$$\lambda_{away} = \frac{\text{Average goals scored by away team} + \text{Average goals conceded by home team}}{2}$$

Replication From Article

For this section, we are replicating a portion of the article. Specifically, the article uses Monte Carlo to predict the results of the match where Manchester City plays Manchester United. We will compare the results we obtain with those from the article.

Manchester City (Home) vs Manchester United (Away) 4/30/2012

Exploratory Plot of Outcomes Before Match



The exploratory plot above, plot A, shows the outcomes of matches for Manchester City (Man City) as the home team in the 2011-2012 season. Specifically, it includes all of Man City's home games leading up to the match that they played against Manchester United (Man United) as the away team on April 30, 2012. The outcomes of the matches are measured as a win (three points), loss (one point) or draw (zero points). Manchester City won the vast majority of those matches and had very few losses. Manchester City had only one draw and no losses that season, leading up to the game we are interested in predicting for.

The exploratory plot above, plot B, shows the outcomes of matches for Manchester United as the away team in the 2011-2012 season leading up to the team's match at Manchester City on April 30, 2012. Manchester United won most of those matches, however, it also had some losses and draws.

From these exploratory plots it seems reasonable to predict that Manchester City would win the match against Manchester United due to the patterns of wins, losses, and draws that each team had during the season, before the match.

Prediction for Match

Table 1. Example Contingency Table for Manchester City.

		Final Rank		
		First	Second	Sum
Match result	Win	3,304,376	1,854,511	5,158,887
	Draw	427,562	1,850,343	2,277,905
	Loss	36,049	2,527,159	2,563,208
	Sum	3,767,987	6,232,013	10,000,000

The article specifically investigates the Manchester City - Manchester United match on April 30, 2012. The paper creates a table of Monte Carlo simulation results, with lambda values based on all of the matches each team played before the match against each other. Ignoring the final rank, this table shows the results of each of the 10,000,000 Monte Carlo simulations in terms of the home team, Manchester City (table shown above and taken from the article by Jiri Lahvicka). The article's Monte Carlo simulation predicted Manchester City winning the match just over 51% of the time.

```
##   Match Result Occurances Percent Article Result
## 1          Win      49059  49.059         51.589
## 2         Loss      32671  32.671         22.779
## 3         Draw      18270  18.270         25.632
```

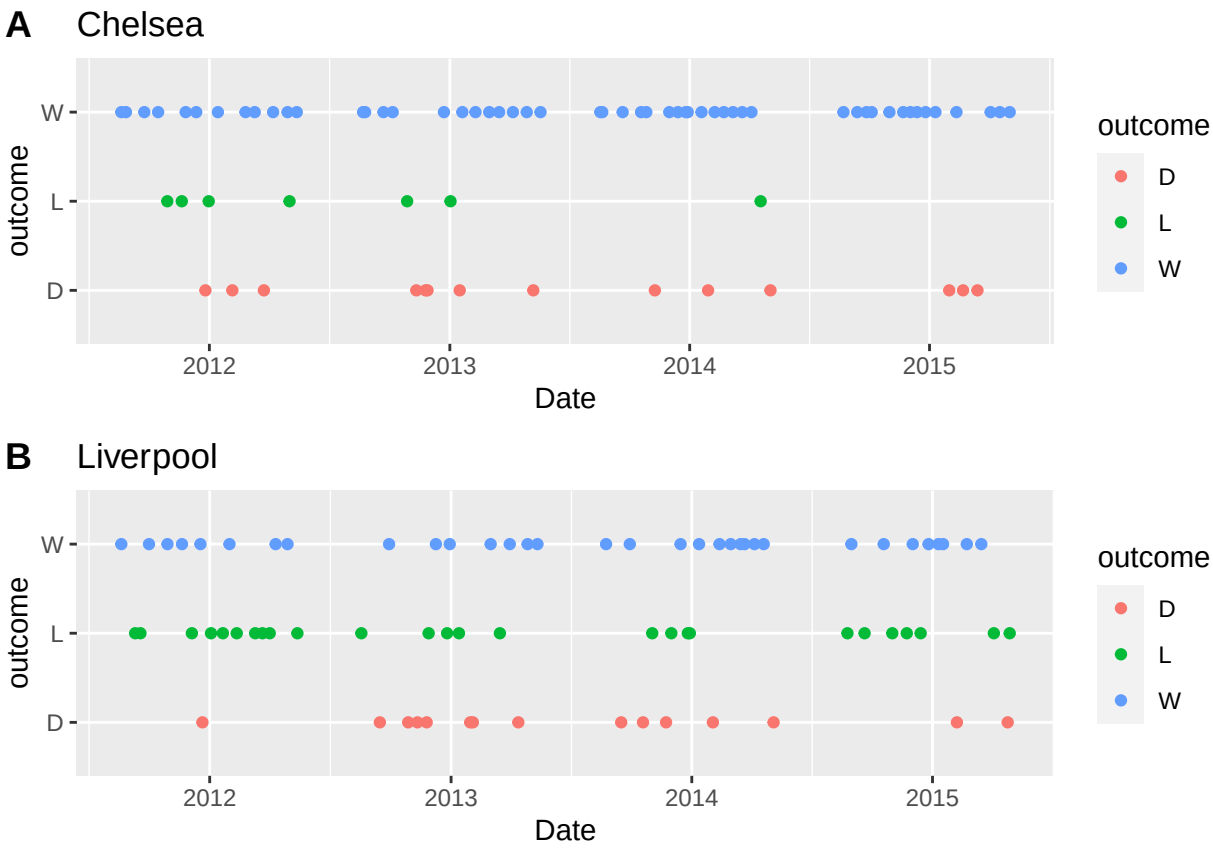
From our Monte Carlo simulations we also created a table in terms of the home team and our Monte Carlo simulation predicted Manchester City would win the match just over 49% of the time. The actual result of the game was that Manchester City did win 1-0. Our result differs slightly from the result in the article and this could be due to the randomness of sampling and our decreased simulation size. We ran 100,000 simulations instead of 10,000,000 because of our limited computational power.

Further Exploration

We decided to go a bit further than the paper, and run simulations to predict the outcomes of other matches. We chose to run simulations for three matches that resulted in either a home draw, home win, or home loss. Something we noticed is that our predictions gave an advantage to the home teams rather than the away teams. This makes sense because generally home teams tend to be more supported and well rested when games are played at their home stadiums. This means that they often play better than the away team.

Chelsea (Home) vs Liverpool (Away) 5/10/2015

Exploratory Plot of Outcomes Before Match



The exploratory plot above, plot A, shows the outcomes of matches for Chelsea as the home team in the seasons from 2011 to 2015 up until the 2014-2015 season match against Liverpool on May 10, 2015. Chelsea won the majority of those matches, had some draws, and few losses.

The exploratory plot above, plot B, shows the outcomes of matches for Liverpool as the away team in the seasons from 2011 to 2015 up until the 2014-2015 season match against Chelsea. Liverpool had much more mixed results than Chelsea, with almost an equal number of losses and wins, as well as some draws. From these exploratory plots it is somewhat difficult to guess which team would have won the match, but Chelsea seems to be a strong guess because the team has had more success than Liverpool.

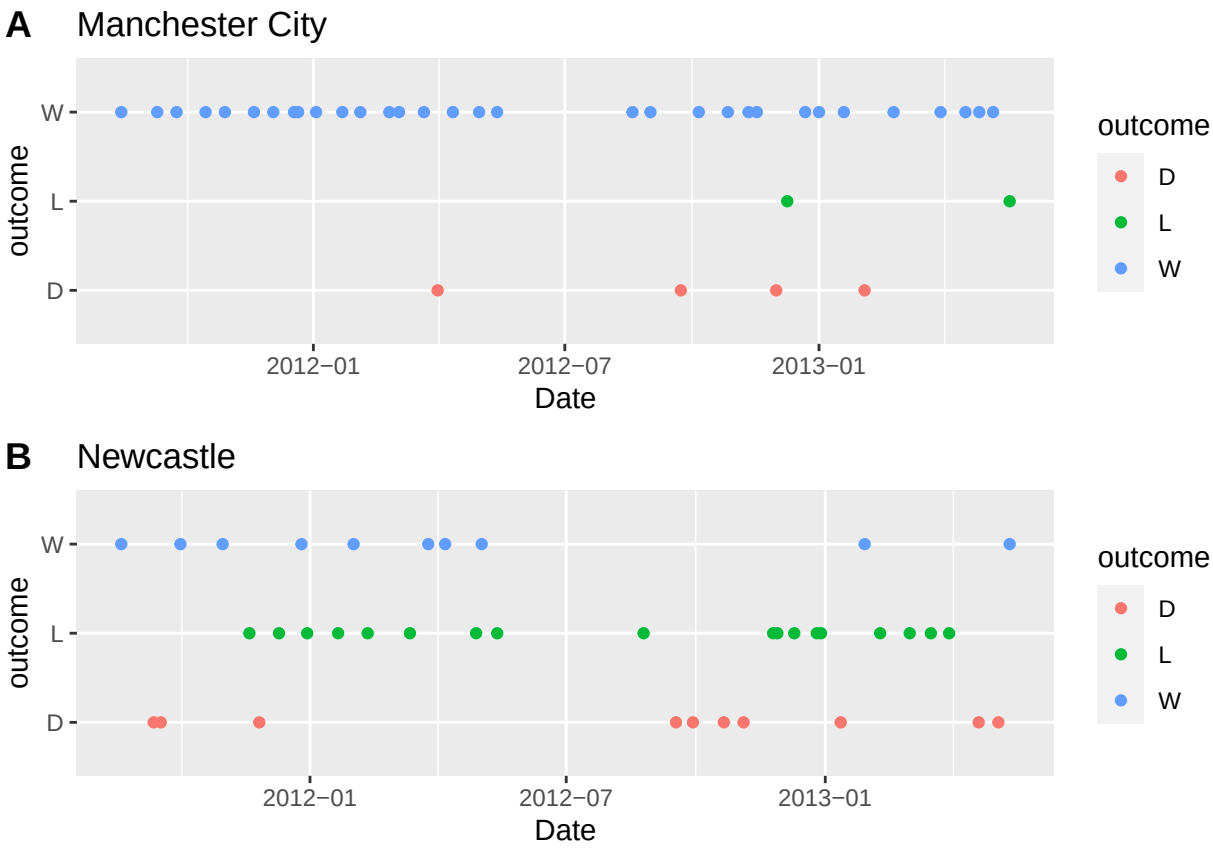
Prediction for Match

```
##   Match Result Occurances Percent
## 1           Win      51999  51.999
## 2          Loss      29180  29.180
## 3          Draw      18821  18.821
```

We ran simulations of matches that modeled Chelsea as the home team and Liverpool as the away team to predict the outcome of the May 10, 2015 match. We compiled our results into a table that models our predictions in terms of the home team (Chelsea). Our simulations resulted in Chelsea winning the match against Liverpool 52% of the time and losing to Liverpool 29% of the time. Based on our simulations, we would have predicted that Chelsea would win the match, however, the game resulted in a draw. Our simulations predicted a draw only about 19% of the time, so this result is fairly surprising.

Manchester City (Home) vs Newcastle (Away) 8/19/2013

Exploratory Plot of Outcomes Before Match



The exploratory plot above, plot A, shows the outcomes of matches for Manchester City as the home team in the seasons from 2011 to 2013 up until the 2013-2014 season match against Newcastle on August 19, 2013. Manchester City won the vast majority of those matches and had very few losses and draws.

The exploratory plot above, plot B, shows the outcomes of matches for Newcastle as the away team in the seasons from 2011 to 2013 up until the 2013-2014 season match against Manchester City. Newcastle won just a few of those matches, with more losses and some draws. From these exploratory plots it seems reasonable to guess that Manchester City would win the match against Newcastle.

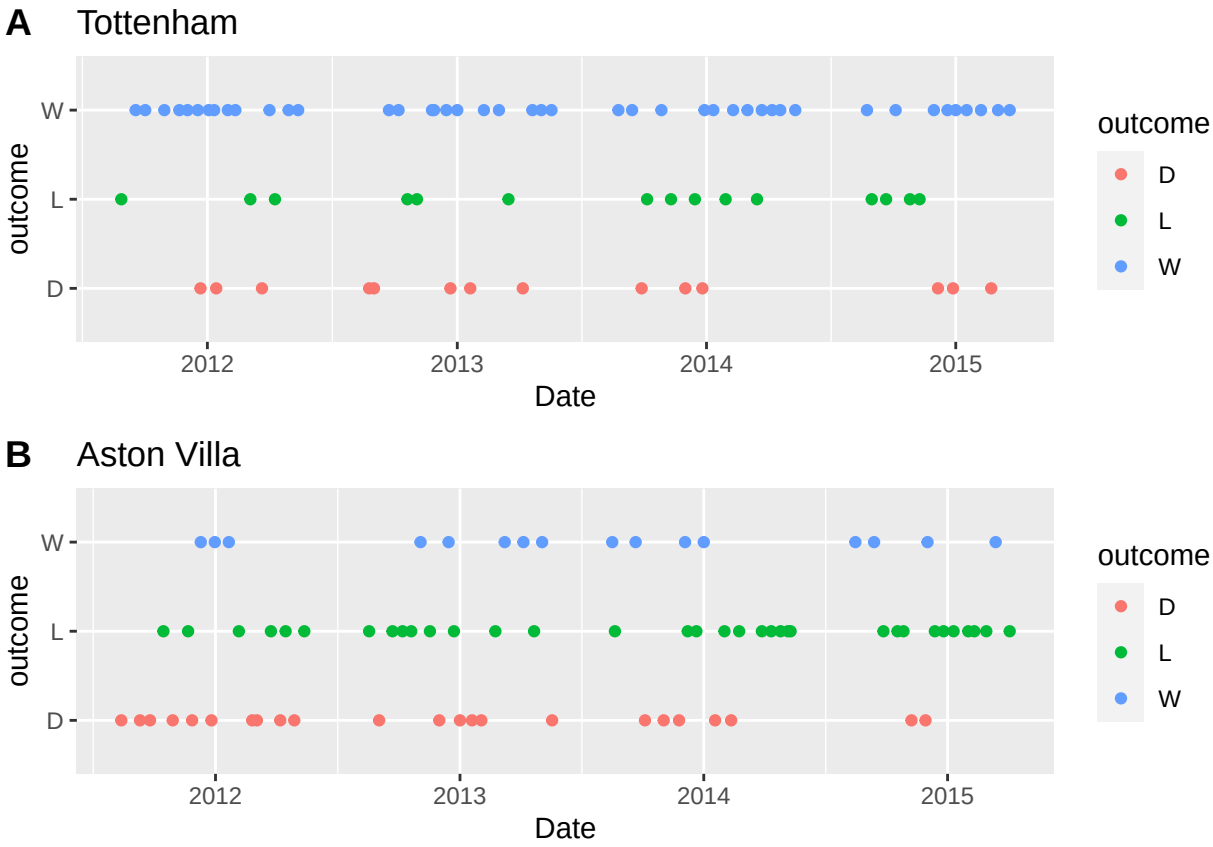
Prediction for Match

```
##   Match Result Occurances Percent
## 1           Win      60084  60.084
## 2          Loss      23188  23.188
## 3          Draw      16728  16.728
```

We ran simulations of matches that modeled Manchester City as the home team and Newcastle as the away team to predict the outcome of the August 19, 2013 match. We compiled our results into a table that models our predictions in terms of the home team (Manchester City). Our simulations resulted in Manchester City winning the match against Newcastle about 60% of the time and losing to Newcastle about 23% of the time. Based on our simulations, we would have predicted Manchester City winning against Newcastle. The simulations resulted in an accurate prediction, as Manchester City beat Newcastle 4-0.

Sunderland (Home) vs Aston Villa (Away) 04/11/2015

Exploratory Plot of Outcomes Before Match



The exploratory plot above, plot A, shows the outcomes of matches for Stoke as the home team in the seasons from 2011 to 2015 up until the 2014-2015 season match against Chelsea on December 22, 2014. Stoke had mixed results of wins and draws, and some losses.

The exploratory plot above, plot B, shows the outcomes of matches for Chelsea as the away team in the seasons from 2011 to 2015 up until the 2014-2015 season match against Stoke. Chelsea had mixed results of wins, losses, and few draws. From these past results, it is hard to predict who would have won the match.

Prediction for Match

```
##   Match Result Occurances Percent
## 1           Win      36912  36.912
## 2          Loss      43949  43.949
## 3          Draw      19139  19.139
```

We ran simulations of matches that modeled Stoke as the home team and Chelsea as the away team to predict the outcome of the December 22, 2014 match. We compiled our results into a table that models our predictions in terms of the home team (Stoke). Our simulations resulted in Stoke winning the match against Chelsea about 37% of the time and losing to Chelsea about 44% of the time. Based on our simulations, we would have predicted Stoke losing the match. This prediction would have been accurate, as Stoke lost the match to Chelsea 0-2.

Estimating Final Season Ranking

Table 1: Ranking Probabilities for Liverpool

Rank.5	Rank.6	Rank.T4	Rank.T5
0.60106	0.13623	0.10592	0.1568

Using the Monte Carlo simulation method that we had used for predicting outcomes of a few matches, the next step was to try and estimate a team's final ranking. The team we chose to do this estimate for was Liverpool. When looking at the actual final rankings of the teams, it seemed uncontested that Chelsea would end up ranking first which is why we ended up choosing Liverpool.

In order to get the final ranking, we first conducted the Monte Carlo simulation on the Chelsea versus Liverpool match we had picked for the last step. We then ran this simulation two more times on the matches Liverpool had left in their season. Once we had these results we could look into the probabilities of each final ranking for Liverpool. In order to figure out the other team's final rankings, we took the results of all of their home and away games and converted them to points depending on that result, 3 for a win, 1 for a draw, and 0 for a loss. Once converted to points, we were able to add up the total amount of points for each team and sort them to see each teams actual final ranking from the 2014-15 season. Since we knew all of the other team's final rankings, we were able to fit Liverpool into the results where they belonged based on the Monte Carlo results.

The table above shows the probability of each possible ranking for Liverpool based on their previous matches in the season and the Monte Carlo simulations. The overall probabilities of all possible scenarios for Liverpool's final three games based on our simulations were calculated. Then based on these probabilities, the total amount of possible points for those last three games was calculated. This total was then added to Liverpool's point total from their previous games in the season. If the point value of the final three games was less than 3, Liverpool would be ranked 6th. If the point value was between 3 and 9, they would be ranked 5th. If the point value was exactly 3, they would end up tying for 5th, and finally, if the point value was exactly 9, they would end up tying for 4th. In the event that two teams tie in total points in the English Premier League, the criteria used to determine final rank next goes to the total goal difference and then total amount of goals scored. Using these criterion, the final rank can be fully determined. In our case, we chose to just leave the results as ties that could be later determined by these other factors. The probabilities of each of these scenarios were calculated and the most likely result was that they would end up ranking 5th with a probability of about 68.719%. In the actual results, they ended up placing 6th, which our simulation said would only happen with a probability of 16.95%.

References

"Data Files: England". *Football-Data.co.uk*, 15 December 2021, http://www.football-data.co.uk/englandm.php

Lahvička, Jiří. "Using Monte Carlo Simulation to Calculate Match Importance: The Case of English Premier League." Journal of Sports Economics, vol. 16, no. 4, May 2015, pp. 390–409, doi:10.1177/1527002513490172.