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 for the Manchester City vs 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 used four seasons ranging from years 2011-2015. Each team played 19 away games and 19 home games. 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.

MC Estimation

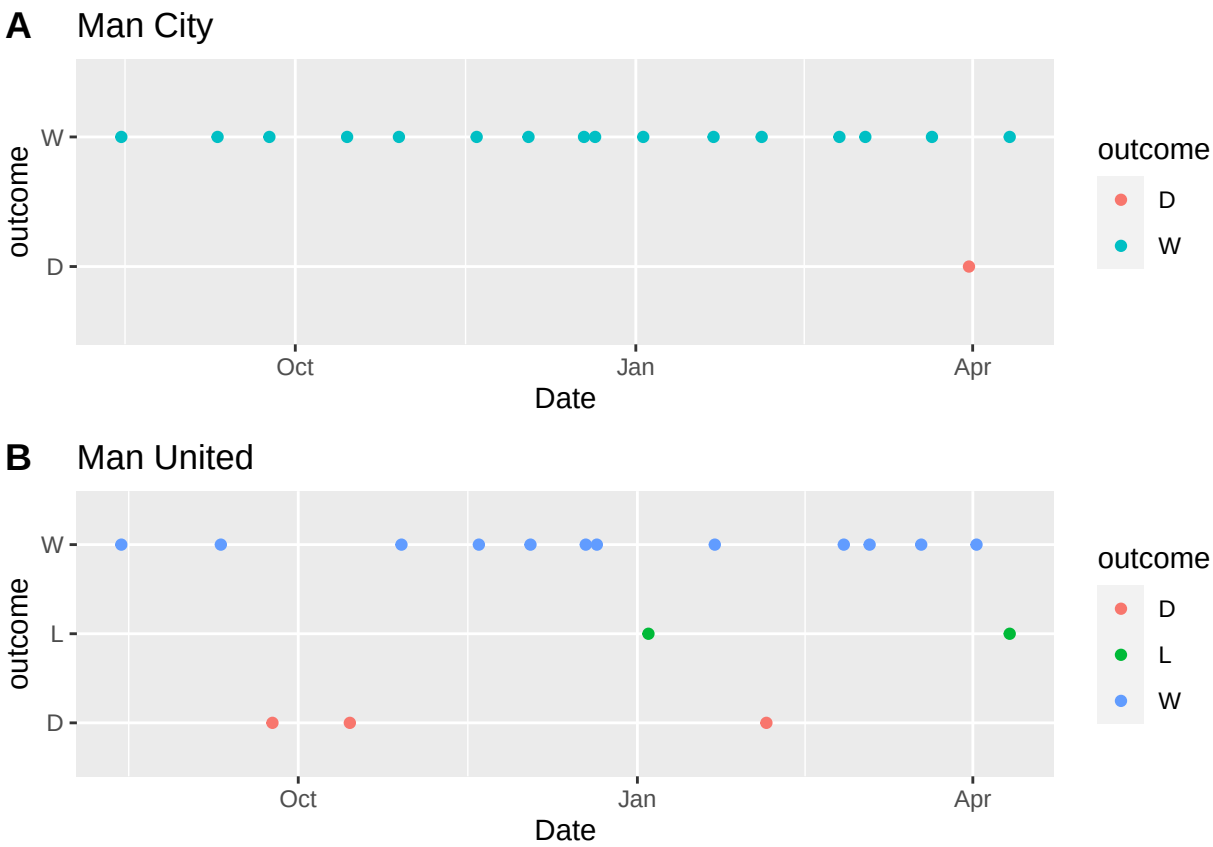
We estimated the lambda home and lambda away values, which are the expected goals scored by the home and away team respectively, using Monte Carlo. These lambda values are assumed to be independent Poisson distributed variables and are calculated using the last 19 matches for each team. The article ran 10,000,000 simulations, but due to our low computational power, we chose to run 100,000 simulations.

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

Replication From Article

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

Exploratory Plot of Outcomes Before Match



Prediction for Match Predictions are in terms of the home team
Actual result was a home win (1-0)

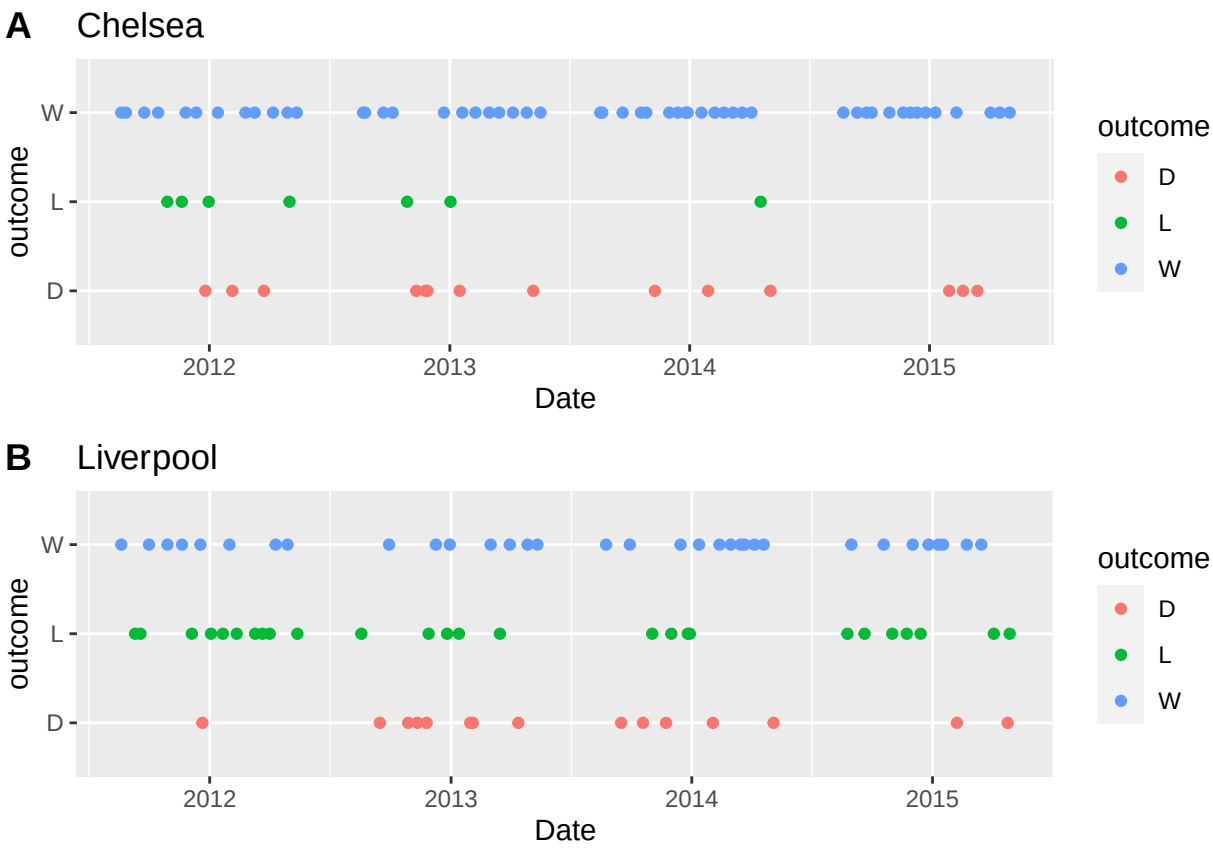
```
##   Match Result Occurances Percent Article Result
## 1           Win      49059  49.059         51.589
## 2          Loss      32671  32.671         22.779
## 3          Draw      18270  18.270         25.632
```

The results of the article differ slightly from the results we found. This is likely due to the randomness of sampling, as well as the increased number of simulations the paper ran.

Further Exploration

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

Exploratory Plot of Outcomes Before Match



Prediction for Match

Predictions are in terms of the home team
Actual result was a draw

```
##   Match Result Occurances Percent
## 1           Win      51734  51.734
## 2          Loss      29241  29.241
## 3          Draw      19025  19.025
```

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

Exploratory Plot of Outcomes Before Match

Prediction for Match

Predictions are in terms of the home team
Actual result was a home win

```
##   Match Result Occurances Percent
## 1           Win      60592  60.592
## 2          Loss      22848  22.848
## 3          Draw      16560  16.560
```

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

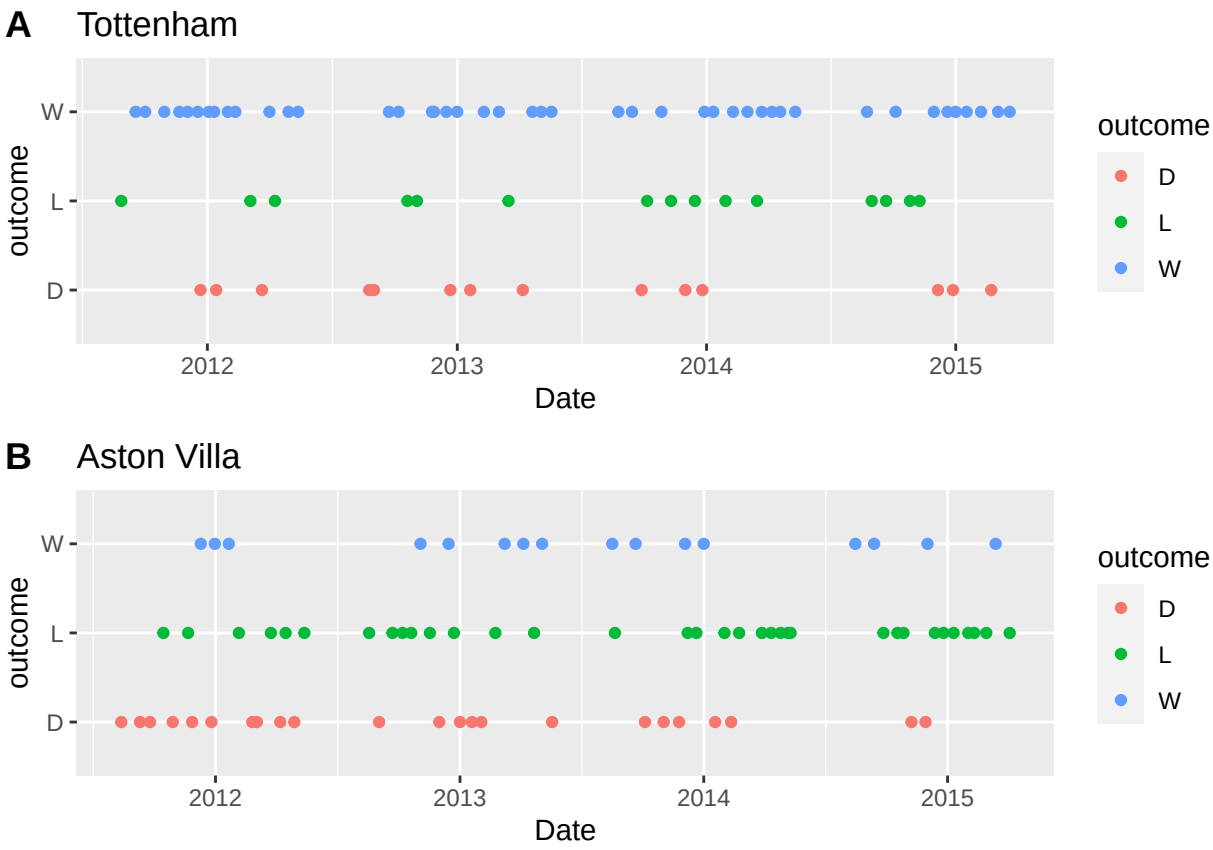
Exploratory Plot of Outcomes Before Match

```
tott <- Home_Function("Tottenham")
tott$Date <- as.Date(tott$Date, "%d/%m/%y")
tott <- tott[1:73,]

av <- Away_Function("Aston Villa")
av$Date <- as.Date(av$Date, "%d/%m/%y")
av <- av[1:73,]

p1 <- ggplot(tott) +
  geom_point(aes(Date, outcome, color=outcome)) + ggtitle("Tottenham")
```

```
p2 <- ggplot(av) +
  geom_point(aes(Date, outcome, color=outcome)) + ggtitle("Aston Villa")
plot_grid(p1, p2, labels="AUTO", ncol=1)
```



Prediction for Match

Predictions are in terms of the home team
Actual result was a home loss

```
##   Match Result Occurances Percent
## 1           Win      53091  53.091
## 2          Loss      28584  28.584
## 3          Draw      18325  18.325
```

Estimating final season ranking