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.

Monte Carlo Estimation

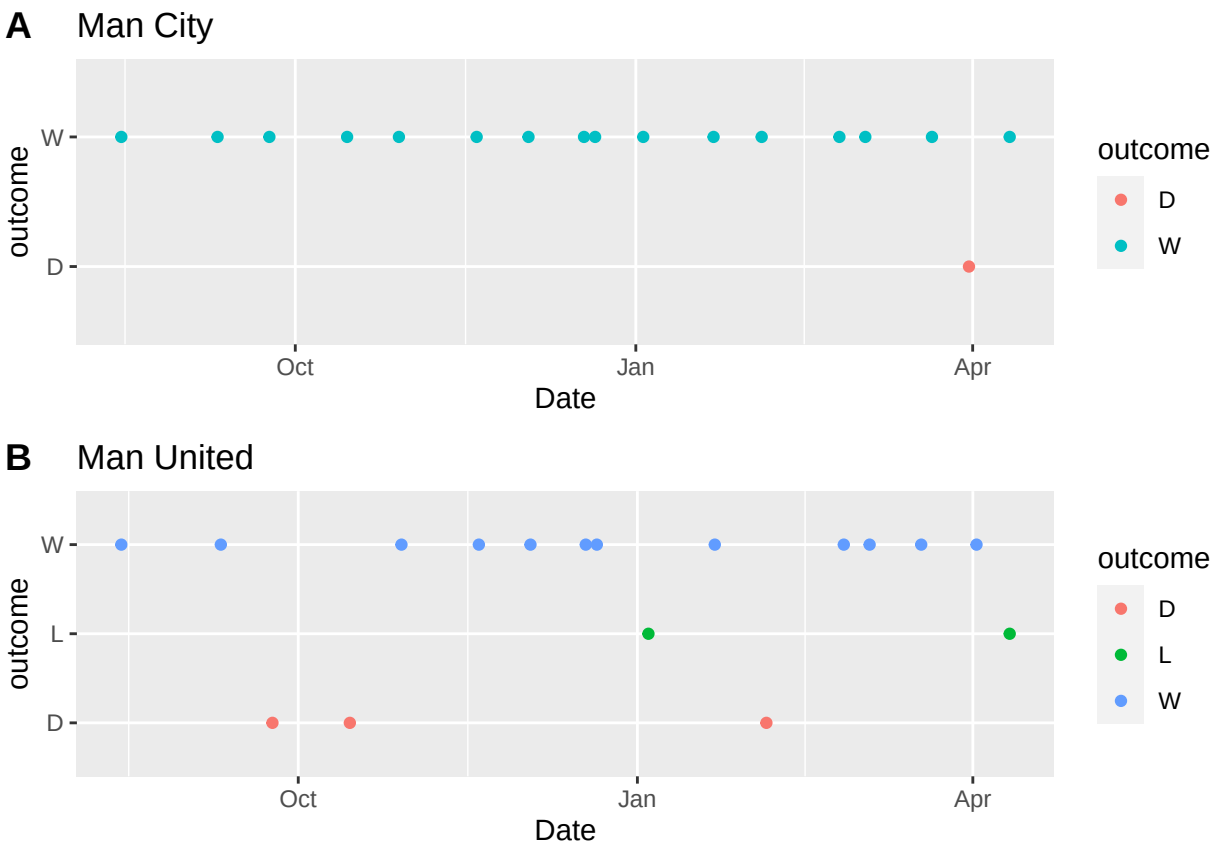
We estimated the lambda home and lambda away values, which are the expected goals scored by the home and away team respectively, using Monte Carlo. These lambda values are assumed to be independent Poisson distributed variables and are calculated using the last 19 matches for each team. The article ran 10,000,000 simulations, but due to our low computational power, we chose to run 100,000 simulations. The general purpose of Monte Carlo is to model the probability of different outcomes and reduce uncertainty. The specific purpose of the paper's Monte Carlo simulations is to

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

Replication From Article

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 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 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. It doesn't appear Man City had any draws. 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. Man United won most of those matches, had some losses and very few draws.

Prediction for Match

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

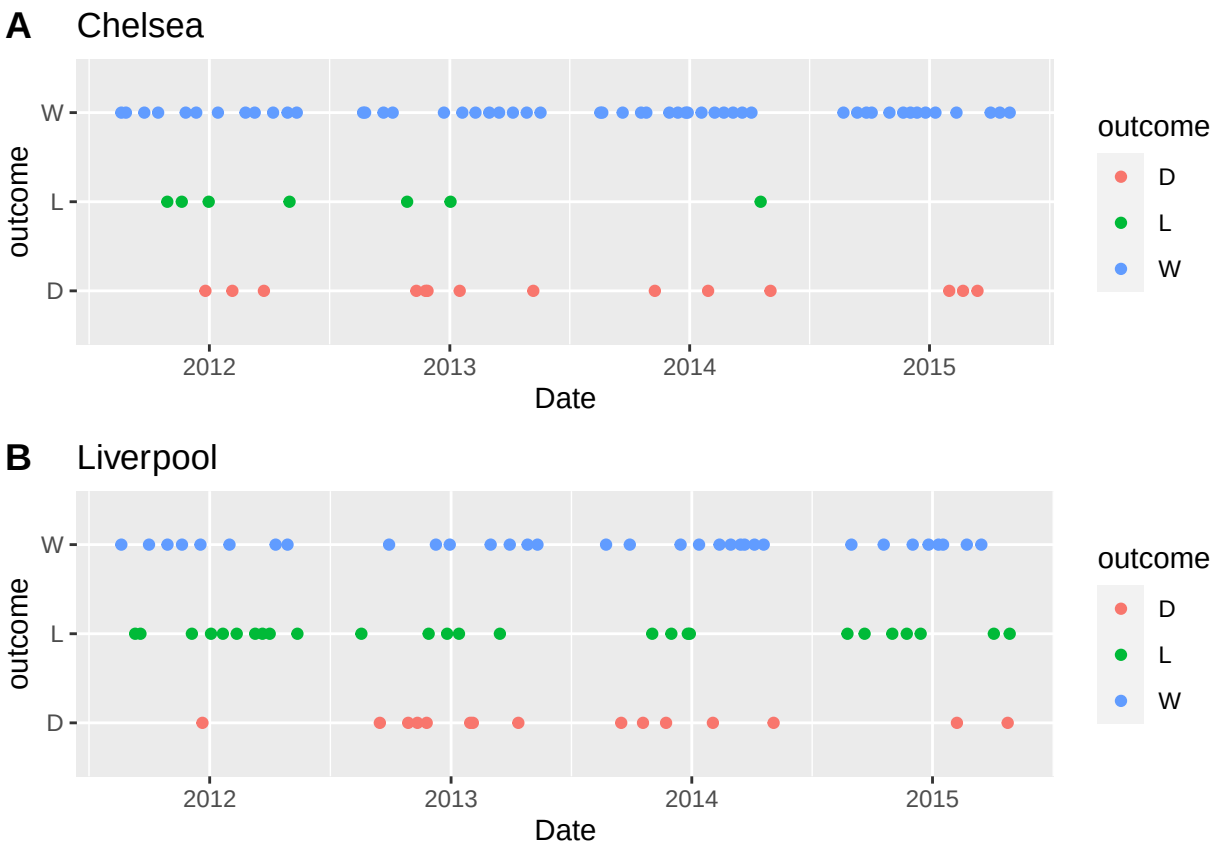
```
##   Match Result Occurances Percent Article Result
## 1           Win      49059  49.059         51.589
## 2          Loss      32671  32.671         22.779
## 3          Draw      18270  18.270         25.632
```

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



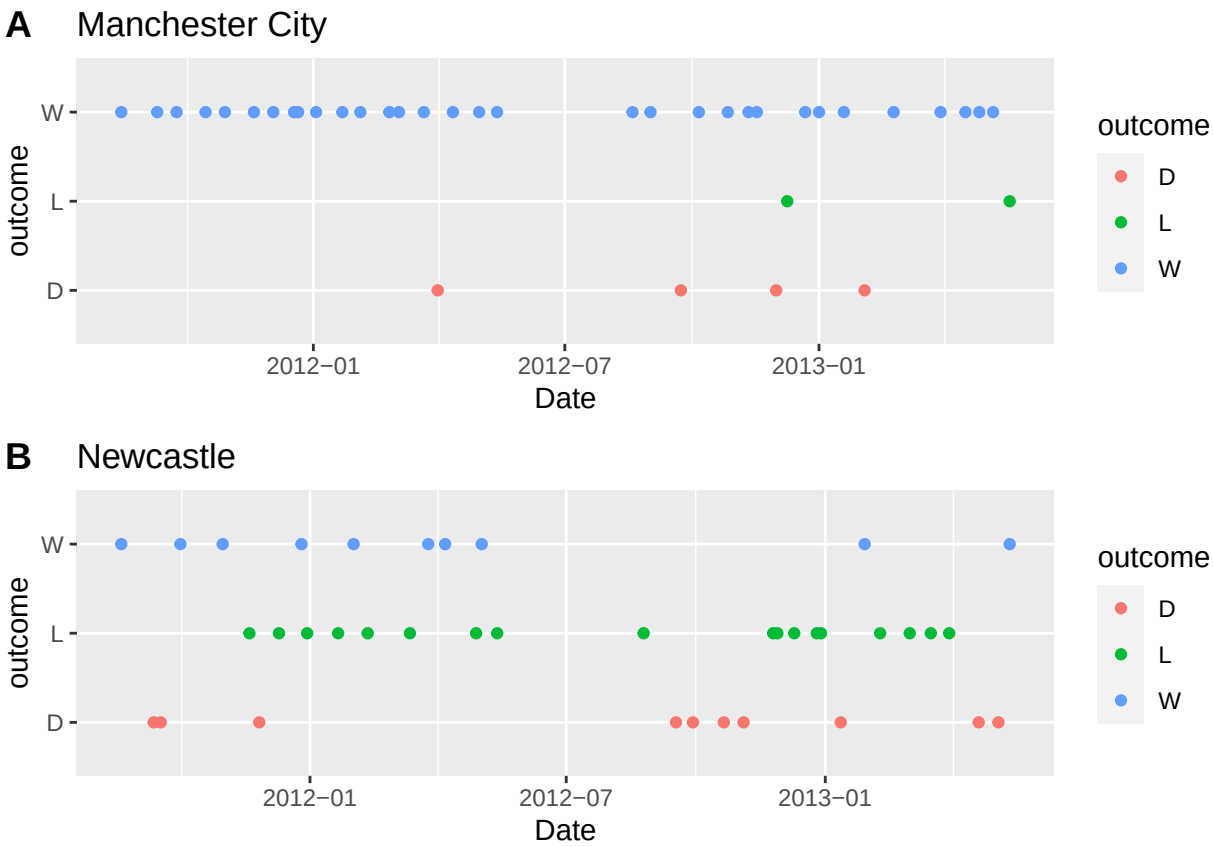
Prediction for Match

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



The exploratory plot above, plot A, shows the outcomes of matches for Manchester City as the home team in the 2013-2014 season. Specifically, it includes all of Man City's home games leading up to the match that they played against Newcastle as the away team on August 19, 2013. Manchester City won the vast majority of those matches and had very few losses and even less draws. The exploratory plot above, plot B, shows the outcomes of matches for Newcastle as the away team in the 2011-2012 season leading up to the team's match at and against Man City on August 19, 2013. Newcastle won most of those matches, had some losses and very few draws.

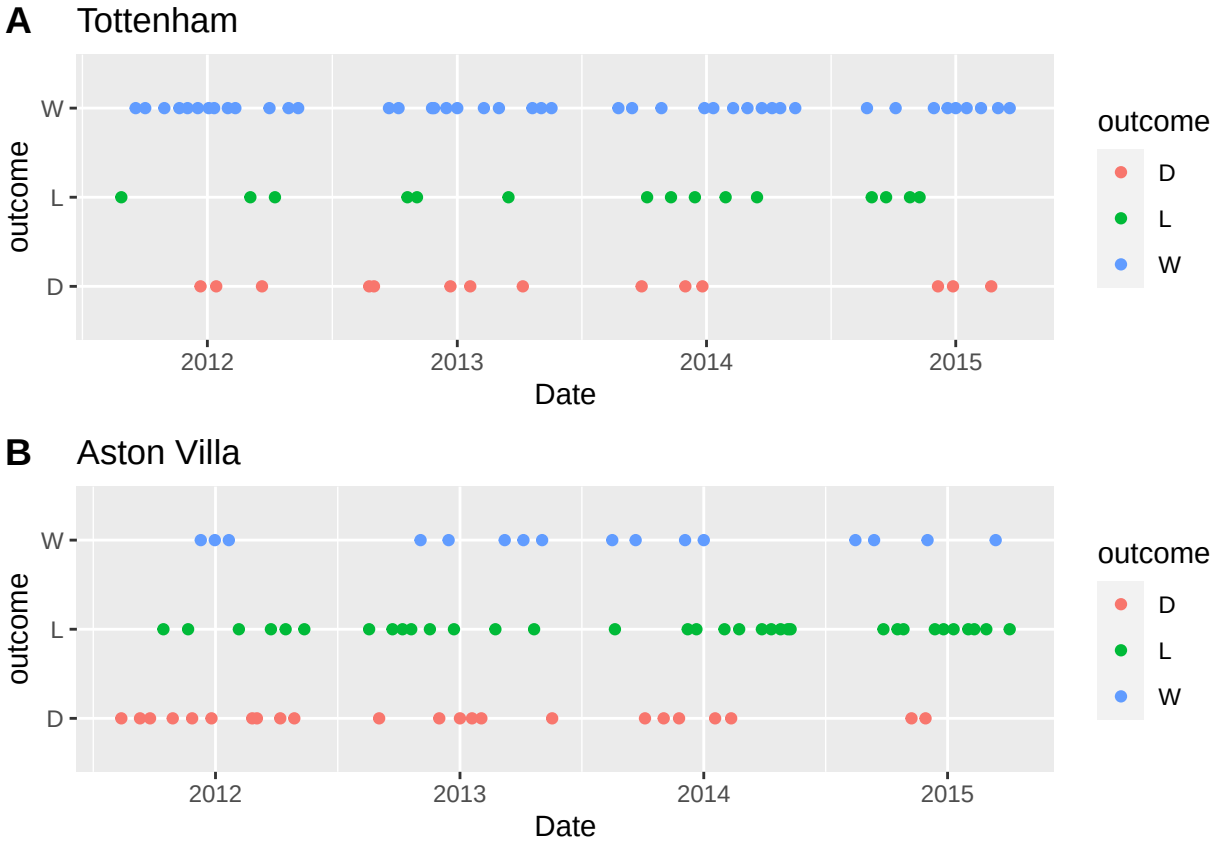
Prediction for Match

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



The exploratory plot above, plot A, shows the outcomes of matches for Tottenham as the home team in the 2014-2015 season. Specifically, it includes all of Tottenham's home games leading up to the match that they played against Aston Villa as the away team on April 11, 2015. Tottenham 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 Aston Villa as the away team in the 2014-2015 season leading up to the team's match at and against Tottenham on April 11, 2015. Aston Villa won some of its matches, but its matches mainly resulted in draws and losses.

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

```
teams1415 <- unique(S14_15$HomeTeam)
S14_15$HomePoints <- rep(NA, length(S14_15$Date))
S14_15$AwayPoints <- rep(NA, length(S14_15$Date))
for(i in 1:length(S14_15$HomePoints)){
```

```
  if(S14_15$FTR[i] == "H"){
    S14_15$HomePoints[i] <- 3
    S14_15$AwayPoints[i] <- 0
  }
  if(S14_15$FTR[i] == "A"){
    S14_15$HomePoints[i] <- 0
    S14_15$AwayPoints[i] <- 3
  }
  if(S14_15$FTR[i] == "D"){
    S14_15$HomePoints[i] <- 1
    S14_15$AwayPoints[i] <- 1
  }
}
head(S14_15)
```

```
##       Date   HomeTeam       AwayTeam FTHG FTAG FTR   Id HomePoints AwayPoints
## 1 16/08/14    Arsenal Crystal Palace    2    1   H 1141          3          0
## 2 16/08/14  Leicester        Everton    2    2   D 1142          1          1
## 3 16/08/14 Man United        Swansea    1    2   A 1143          0          3
## 4 16/08/14        QPR           Hull    0    1   A 1144          0          3
## 5 16/08/14      Stoke    Aston Villa    0    1   A 1145          0          3
## 6 16/08/14  West Brom     Sunderland    2    2   D 1146          1          1
```

```
hpts <- S14_15 %>%
  group_by(HomeTeam) %>%
  summarise(HomePoints = sum(HomePoints))

apts <- S14_15 %>%
  group_by(AwayTeam) %>%
  summarise(AwayPoints = sum(AwayPoints))

tpts <- cbind(apts, hpts)

tpts$TotalPoints <- tpts$AwayPoints + tpts$HomePoints
tpts
```

```
##          AwayTeam AwayPoints       HomeTeam HomePoints TotalPoints
## 1         Arsenal         34        Arsenal         41          75
## 2     Aston Villa         17    Aston Villa         21          38
## 3         Burnley         14        Burnley         19          33
## 4         Chelsea         38        Chelsea         49          87
## 5  Crystal Palace         27 Crystal Palace         21          48
## 6         Everton         19        Everton         28          47
## 7            Hull         15           Hull         20          35
## 8       Leicester         15      Leicester         26          41
## 9       Liverpool         27      Liverpool         35          62
## 10       Man City         34       Man City         45          79
## 11     Man United         26     Man United         44          70
## 12      Newcastle         13      Newcastle         26          39
## 13            QPR          7            QPR         23          30
## 14    Southampton         23    Southampton         37          60
## 15          Stoke         21          Stoke         33          54
## 16     Sunderland         18     Sunderland         20          38
## 17        Swansea         24        Swansea         32          56
```

```
## 18      Tottenham         31      Tottenham         33          64
## 19      West Brom         19      West Brom         25          44
## 20       West Ham         16       West Ham         31          47
```

```
tpts <- tpts[order(-tpts$TotalPoints),]
```

```
#match 1501
Match1501 <- mc_funct(match_id = 1501, home_team = "Liverpool", away_team = "Crystal Palace")
res1501<- table(t(Match1501$ResultH))
res1501 <- as.data.frame(res1501)
res1501$Percent <- res1501$Freq/100000 * 100
order <- c("Win", "Loss", "Draw")
res1501 <- res1501 %>%
  slice(match(Var1, order))
colnames(res1501) <- c("Match Result", "Occurances", "Percent")
res1501
```

```
##   Match Result Occurances Percent
## 1          Win      45592  45.592
## 2         Loss      31690  31.690
## 3         Draw      22718  22.718
```

```
#match 1520
Match1520 <- mc_funct(match_id = 1520, home_team = "Stoke", away_team = "Liverpool")
res1520<- table(t(Match1520$ResultH))
res1520 <- as.data.frame(res1520)
res1520$Percent <- res1520$Freq/100000 * 100
order <- c("Win", "Loss", "Draw")
res1520 <- res1520 %>%
  slice(match(Var1, order))
colnames(res1520) <- c("Match Result", "Occurances", "Percent")
res1520
```

```
##   Match Result Occurances Percent
## 1          Win      44826  44.826
## 2         Loss      36197  36.197
## 3         Draw      18977  18.977
```

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.