Effect of rest on soccer and tennis match outcomes

Chloé Lepert

8/9/2018

1. Introduction

Many factors influence the athletic performance in competitions. Most models aiming to predict results focus on the strength of the team and location of the competition, but other factors could have an impact. In this paper we will look at how rest impacts the performance of soccer teams and tennis players.

1.1 Question

We aim to quantify the effect of rest on Football match outcomes. We ask two questions looking at short and long term rest levels of a team:

1. How does the number of days since a previous match impact the performance of a team?
2. How the the game load of a team in the past month, 2 month, and 3 month impact the performance of a team?

We will look at different models in which performance will either be the probability of winning, drawing, and losing or the expected number of goals scored.

1.2 English Football

Before explaining our motivation we will first given an overview of the English and European soccer system so that the reader may better understand our motivation.

English soccer is comprised of multiple competitions.

- Premier league - Top 20 teams in England. They will play each other time once at home and once away, totally 380 games. A win will get a team 3 points, a draw one point, and loss 0 points. At the end of the season points are tallied, and the team with the most points win. The bottom two teams will be relegated to the EFL Championship. The top two teams from EFL Championship will advance to the Premier league. The third to last team in the Premier league and third team in the EFL Championship will play off to play in the Premier league. [**?**]
- FA cup - Open to teams in 10 levels of English football, this a knockout tournament. Teams enter the tournament at different times depending on the level they play in. Teams that play in the Premier league will enter the 3rd round in January with the final in May. [**?**]
- EFL cup - similar to the FA cup but open only to levels 1 through 4. Premier league teams will enter in August with the final in February. Teams playing in European competitions will not play this cup [**?**]
- Champions league - A competition played with team from different European leagues. It starts with a group stage of 32 teams. Leading to 16 teams going though a tournament. Teams not advancing to the tournament are transferred to the Europa league. In the tournament stage teams meet twice once home once away. The return leg can go overtime to decide who gets to move to the next round. The final is played in one game in a location determined ahead of time. [**?**] In England the top three teams in the premier league automatically qualify for the Champions league. The fourth place team qualifies for a play-off round. [**?**]

- Europa league - Similar to the Champions league in format. [?] The 5th place Premier league team qualifies for the group stage, with the winners of the FA cup and EFL cup qualifying for earlier rounds. [?]

International break - depending on the year, the schedule will be inter-spaced with 2-week international breaks, in which players are called to represent their national teams in international friendlies or qualifying matches for continent or world cups.

1.2 Relevance

Football game scheduling varies from league to league. In particular, some leagues take a winter break while others do not. Some speculate that this winter break influences performance in inter-league competitions such as the Champions and Europa league. [?]

Further Champions and Europa league games are played Tuesday through Thursday, while club games are played Friday through Mondays, leading to a variety of rest days going into Champions and Europa league games. (NEED CITATION) Some leagues are more willing to accommodate teams playing in European competitions than others.

1.3 Approach

Due to the few numbers of European competition games played, modeling outcomes of European matches is difficult. We will instead model the outcome of Premier League matches to evaluate the effect of game load and rest.

This approach assumes that European competition games are similar Premier League games. This is mostly true except for the fact that European competition games can result in overtime and are higher stakes; a loss results in elimination. We assume that if we see rest impact outcomes in premier league games, rest is likely to have an impact in European games.

1.4 Literature review

1.4.1 Rest time

Most of research on impacts of game scheduling focuses on injury and measures of certain activities in games (distance ran, number of sprints, etc.), but not on overall game outcomes.

Over the past decades, Carlos Lago Penas conducted a series of study have looking at how physical behaviors of players such as distance run at certain speeds evolve over a series of consecutive games with small rest periods and found minor to no differences in physical behaviors across games. [?, ?]

In 2010, Dupont et al. found a higher injury rates for football players who played two matches a week compared to players playing 1 match a week.

1.4.2 Modeling match outcomes

In 1982, M. J. Maher introduced 2 independent Poisson distributions as a way to model soccer scores. He proposed using a team's attacking strength and its oponent's defensive weakness as predictors of number of goals scores. Maher also found that using a Bivariate Poisson distribution with a correlation of 0.2 improved his model's fit. [?]

In 2003, Karlis and Ntzoufras proposed a diagonal inflated bivariate Poisson model in which the probabilities of draws are increased. [?] They also created an R package to fit biavariate Poisson GLMs which we will utilize. [?]

2 Methods

2.1 Poisson model

The Poisson model is most commonly used to model soccer scores. It assumes that the number of goals scored by a team in a soccer game follows a Poisson distribution of some parameter. We will use a GLM with a poisson link to predict the parameter.

2.1.1 Model

- Let $j = h, a$ be an indicator for whether we are modeling home or away games.
- Let $G_{j,i}$ be the number of home or away goals in game i.
- Let x_i be the predictors for game i.
- Assume that $G_j \sim Poisson(\lambda_j)$. $P_P(G_{j,i} = g_{j,i}) = \frac{\lambda_j^{g_{j,i}} e^{\lambda_{j,i}}}{g_{j,i}!}$
- The parameter λ_j is a linear combination of the predictors X_j: $\lambda_j = X_j \beta_j$

2.1.2 Verifying the Poisson assumption under the null

One of the defining features of the Poisson model is the fact that the mean of a Poisson distribution equals it's variance. As the figure below shows for most teams this equality holds.

REMAKE FIGURE

Results

Simple Poisson GLM

Table ~1 summarizes the GLM models for the number of home and away goals. Rest comes in the model either as number of days since the previous game (1) and (3) or as whether or not the team had more than 6 days of rest (2) and (4). In neither of the 4 models do we see rest as having an effect on the number of goals scored.

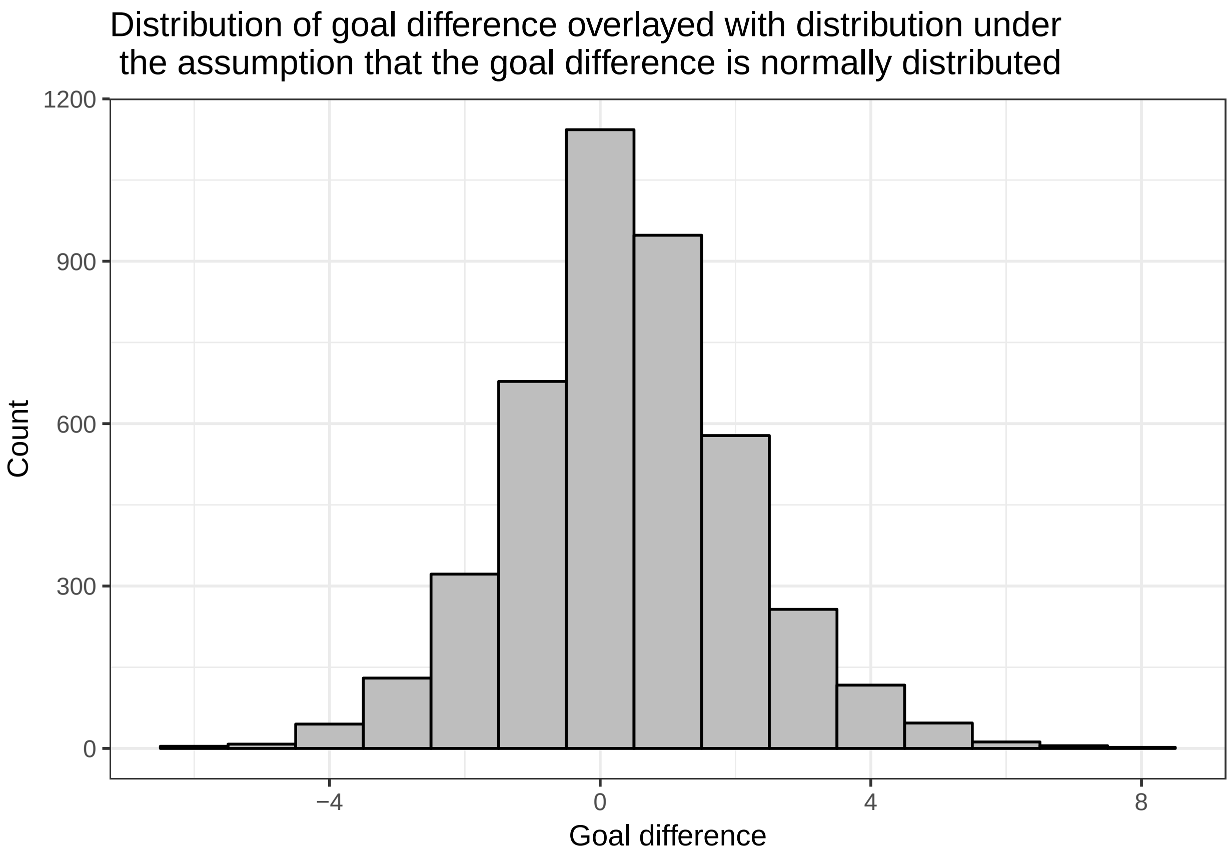
Table 1: Generalized Linear Models with Poisson link

	Dependent variable:			
	hgoal		vgoal	
	(1)	(2)	(3)	(4)
Team_rest	−0.004		−0.007	
	(0.009)		(0.011)	
Opp_rest	−0.001		0.005	
	(0.009)		(0.011)	
Team_rest_bin		−0.007		−0.029
		(0.028)		(0.032)
Opp_rest_bin		−0.019		0.013
		(0.027)		(0.032)
Team_att_str	0.371***	0.372***	0.427***	0.427***
	(0.050)	(0.050)	(0.053)	(0.053)
Opp_def_weak	0.274***	0.277***	0.238***	0.239***
	(0.061)	(0.061)	(0.061)	(0.061)
Team_load	0.015***	0.016***	0.011***	0.011***
	(0.003)	(0.003)	(0.003)	(0.003)
Opp_load	−0.020***	−0.021***	−0.024***	−0.024***
	(0.003)	(0.003)	(0.004)	(0.003)
Constant	−0.036	−0.044	−0.025	−0.021
	(0.205)	(0.198)	(0.231)	(0.223)
Observations	4,296	4,296	4,296	4,296
Log Likelihood	−6,597.132	−6,596.946	−5,804.871	−5,804.693
Akaike Inf. Crit.	13,208.260	13,207.890	11,623.740	11,623.390
Note:				*p<0.1; **p<0.05; ***p<0.01

Linear model for goal difference

```
p <- readRDS("~/Documents/masters_paper/Figures/goalDiffNormal.rds")
p
```

```
## Warning: Computation failed in `stat_function()`:
## object of type 'closure' is not subsettable
```



From lmGoalDiff.R

Discusion

Conclusion

Appendix

References

Notes

- in 1.1 Do we still care about long term?
- 2.1.2 Remake the figure
- Fix all the numbering

Table 2: Linear model for the difference in goals scored

	Dependent variable:
	goal_diff
h_rest	−0.003
	(0.018)
v_rest	−0.002
	(0.018)
h_att_str	0.807***
	(0.105)
h_def_weak	−0.529***
	(0.105)
v_att_str	−0.672***
	(0.092)
v_def_weak	0.795***
	(0.120)
h_load	0.036***
	(0.006)
v_load	−0.026***
	(0.006)
Constant	−0.414
	(0.450)
Observations	4,296
R^2	0.146
Adjusted R^2	0.144
Residual Std. Error	1.604 (df = 4287)
F Statistic	91.666*** (df = 8; 4287)
Note:	*p<0.1; **p<0.05; ***p<0.01