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. [**?**]

1.5 Data

1.5.1 Data set

The R package "engsoccerdata" provides us will all score from matches in England as well as European competition played by English teams over the past 100 years. [**?**] We limit ourselves to matches occurring in seasons 1995 through 2015. The design of English Football stays consistent throughout this time range and our data set is complete for these years.
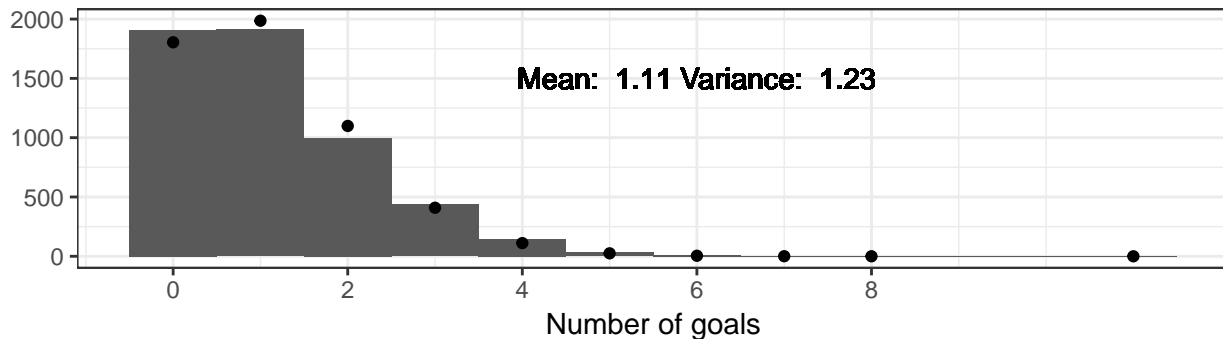
1.5.2 Response variable

We use two different type of response variables: 1. The number of home and away goals and 2. Weather the a team won, tied, or lost. The distribution of home and away goals is shown bellow.

stribution of goals overlayed with expected number of goals assuming a poisson distributi
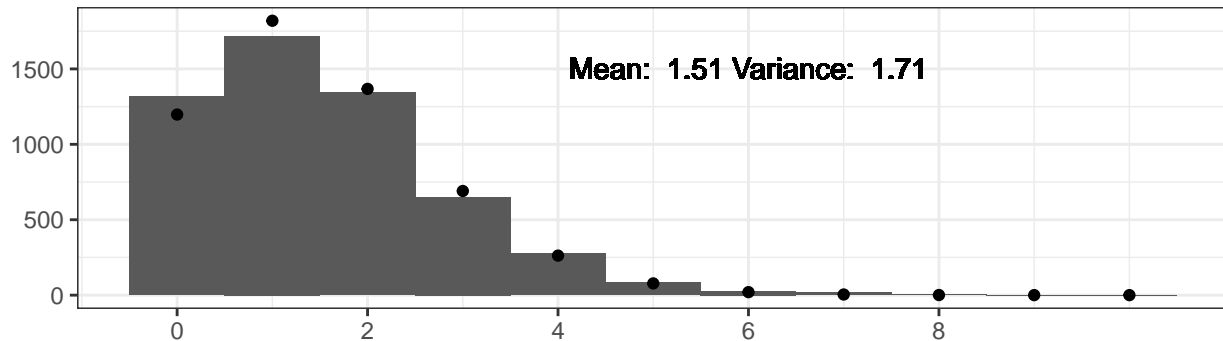

```
## TableGrob (3 x 1) "arrange": 3 grobs
##   z     cells    name                grob
## 1 1 (2-2,1-1) arrange     gtable[layout]
## 2 2 (3-3,1-1) arrange     gtable[layout]
## 3 3 (1-1,1-1) arrange text[GRID.text.88]
```

Most models modeling number of goals in a game assume he number of home and away goals in a game follow an approximately poisson distribution. We see that this approximately true. There tends more games than expected with zero goals and the variance is slightly larger than the mean number of goals.

A game is a home win if the number of home goals exceeds the number of visitor goals, tied if both teams score the same number of goals, and a visitor win if the number of visitor goals exceeds the number of home goals. The distribution of outcomes is shown below.

Table 1: Distribution of game outcomes

	Outcome	Number of games	Share of games
1	Home win	$1,966$	0.458
2	Tie	$1,143$	0.266
3	Visitor win	$1,187$	0.276

We also control for the number of games a team plays per season. Better teams will last longer in in playoff competitions (FA Cup, Europa/Champions League, etc.) that happen concurrently with premier league games. These teams will play more games per season and thus on average have lower rest times.

Distribution of the number of games played by a teams over one season
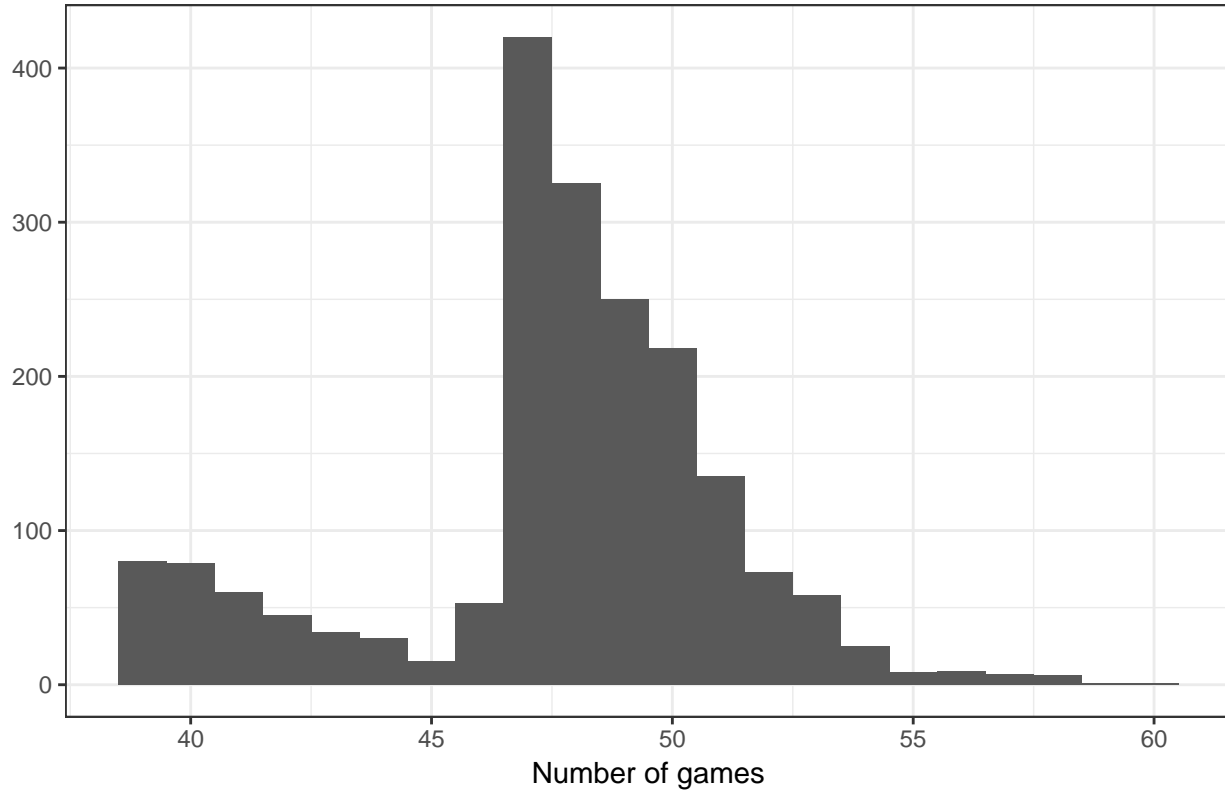
1.5.3 Predictors

While we use all games in the data set to calculate game load and rest time we limit ourselves to games in the premier league in evaluating the effect of rest on outcomes. We do this because design of the premier league in which pair of team plays each other once at home and once away allows for cleaner statistical analysis.

We use this data set to calculate the number of days since the previous game. Before performing out analysis on the effect of rest time we remove games in which either of the team's previous game was over 8 days ago. If team's last non-international game was more than 8 days prior, it is possible that an international break previous happened which we have no way of controlling for.

The distribution of rest time for the home and away team is shown below.

Table 2: Distribution of rest time (in days) for home and away teams

		Visitor						
		2	3	4	5	6	7	8
Home	2	128	9	1	0	3	0	1
	3	12	343	155	22	38	121	7
	4	0	165	238	43	38	91	61
	5	1	26	50	63	31	90	31
	6	3	22	34	42	110	278	41
	7	0	118	74	93	302	798	110
	8	1	3	47	42	32	106	272

We also create a binary variable for weather or not a team is "more" (7-8 days) or "less" (1-6 days) rested and find the following distribution of rest.

Table 3: Distribution of rest time for home and away teams

		Away	
		1-6 days	7-8 days
Home	1-6 days	$1,577$	721
	7-8 days	712	$1,286$

1.5.4 Control variables

Using Premier games we calculate a teams attacking and defensive weakness. [**?**] The attacking strength is the ratio of the average number of goals score by a team in a season to the average number of goals scored in the league that season. The defensive weakness is the ratio of the average number of goal conceded by a team in a season to the average number of goals concede in the league that season. For each team the attacking strength is computed for home and away games. The better a team the higher its attacking strength and the lower its defensive weakness strength will be. We will use the previous years attacking and defensive weakness to control for a team's ability. We exclude recently promoted teams from our analysis as we cannot use their previous season to calculate such strengths. The distribution of attacking and defensive strengths is shown below.

Distribution of attacking strength and defensive weakness for home and away
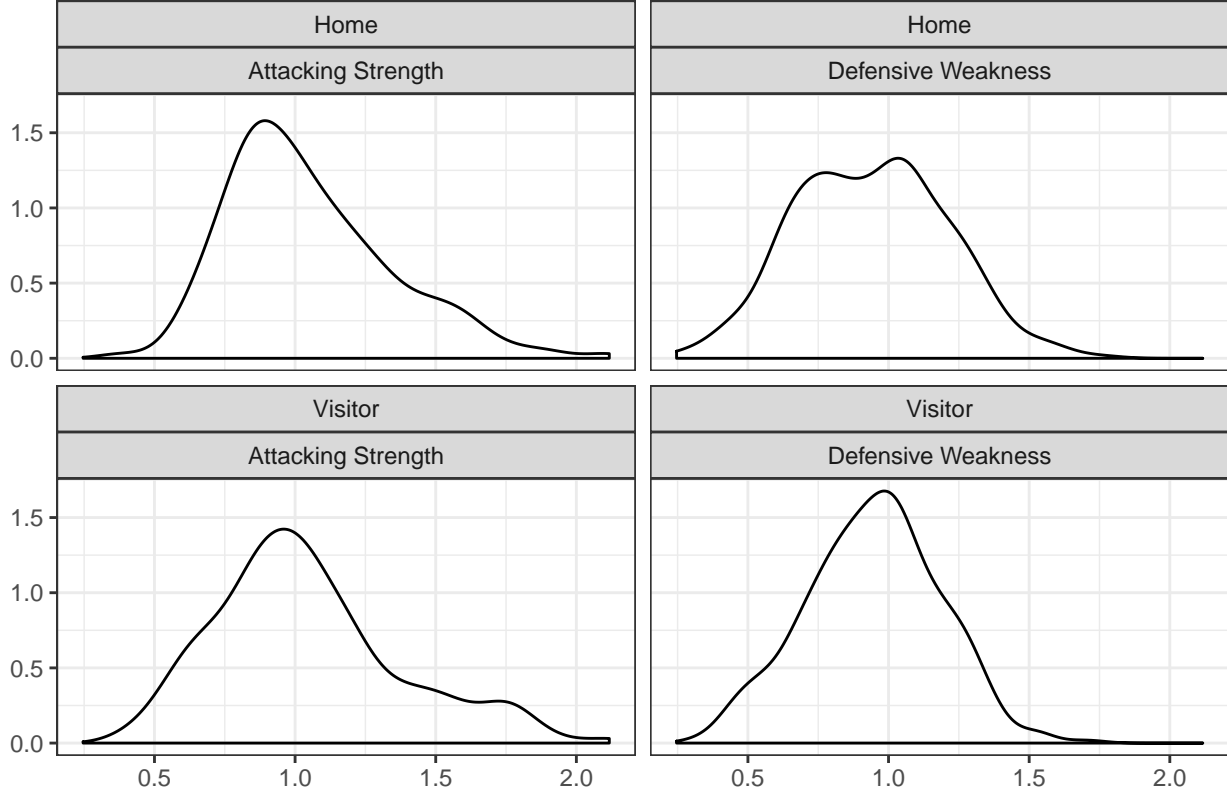
1.5.5 Further subsetting

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.


Variance and Mean of the number of goals in a game split by home and away
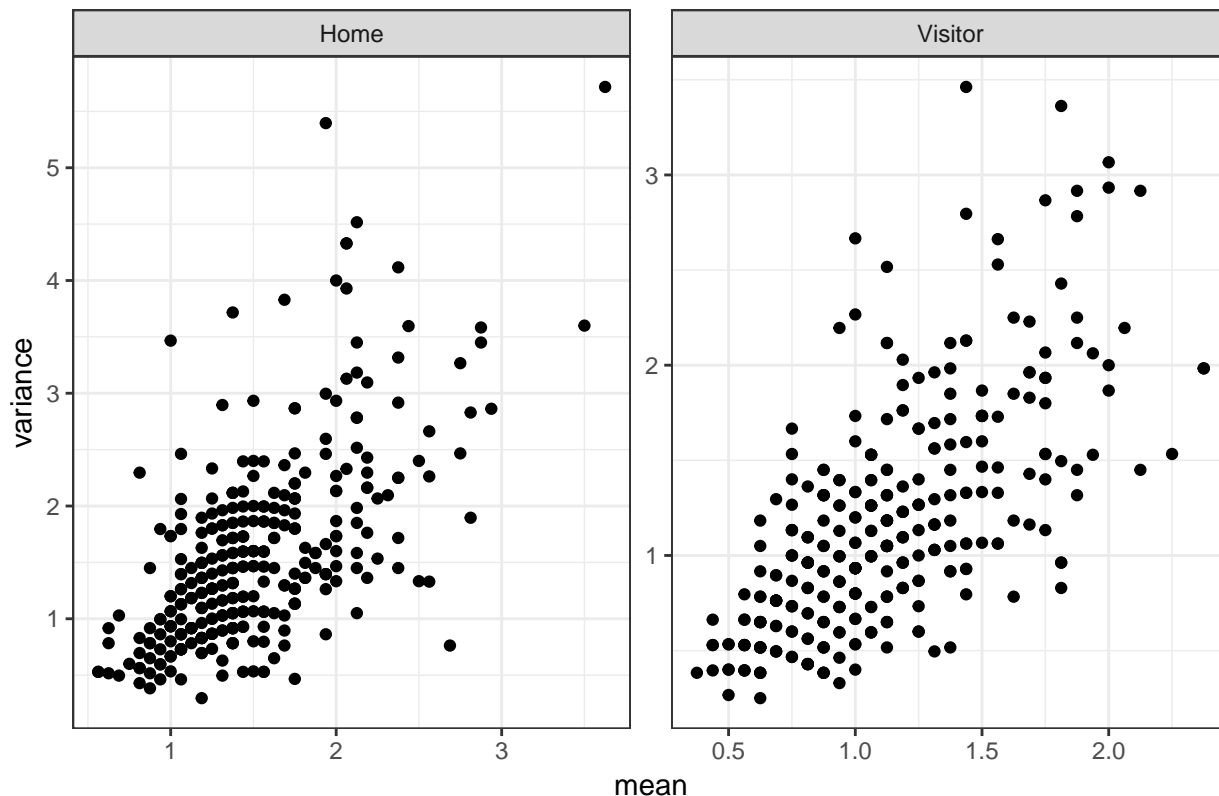
2.2 Linear model for goal difference

The square root of a poisson distribution can be approximated by a nornal distribution (cite). We can therefore approximate the difference in square root goals as the difference of two normal distributions, which is itself a normal distribution. This allows us to fit a normal model to the transformed goal difference. The figure below shows the distribution of normal goal difference and its approximation as a normal distribution.

Distribution of square root goal difference overlayed with distribution under the assumption that the goal difference is normally distributed
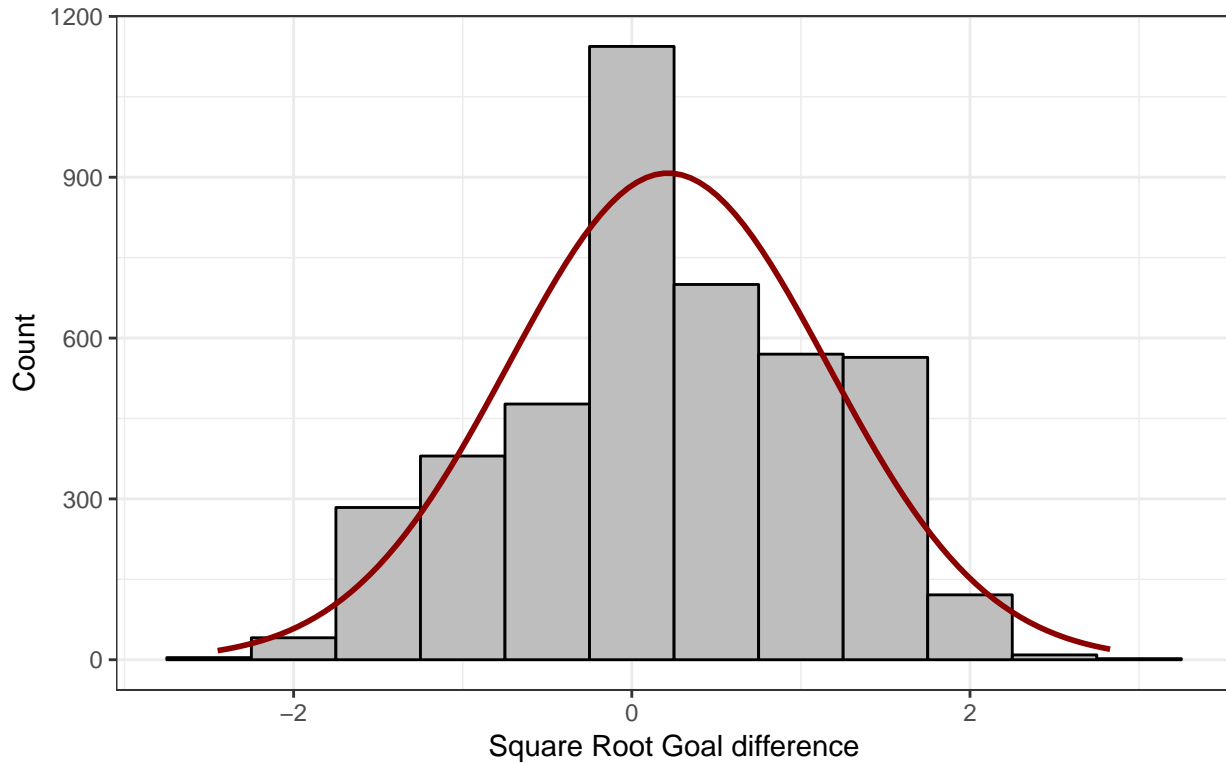
Our model aims to find a linear combination of our predictors that best approximates the square root goal difference

$$\sqrt{G_{h,i}} - \sqrt{G_{v,i}} = X_j\beta_j$$

In this model we could use difference in the number of rest days as a predictor instead of number of rest days for the home team and number of rest days for the away team. Using the difference in rest as a predictor is equivalent to using the the rest days for the home team and away teams and forcing their coefficients in the model to be the opposite. We chose to not add this constraints and allow home and away rest to vary freely.

Results

Simple Poisson GLM

Table 4 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.

The strongest effects come from the attacking strength of the home team and defensive weakness of the visitor team. This not to surprising as most models looking to predict soccer outcomes use the product of these two quantities as the expected number of goals. A team with a high attacking strength should score more since it has in the past, and a team with a high defensive weakness should take in more goals as it has in the past.

We also see an effect from the game load of a team. Game load also measure the quality of a team. Better teams will qualify for more games and thus have a higher game load. We expect to see an increase in a team's game load increase its predicted number of goals and an increase in its opponent's game load decrease a team's predicted number of goal; which is what we see.

Table 4: 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

Table 5 summarizes the model for the square root goal difference. We find that the rest of the home and away team have no significant impact on the goal difference. Predictors highlighting the strength of the teams involved are most important.

Table 5: Linear model for the difference in goals scored

	Dependent variable:
	goal_diff
h_rest	−0.005
	(0.010)
v_rest	−0.003
	(0.010)
h_att_str	0.399***
	(0.057)
h_def_weak	−0.277***
	(0.057)
v_att_str	−0.380***
	(0.050)
v_def_weak	0.364***
	(0.066)
h_load	0.019***
	(0.003)
v_load	−0.015***
	(0.003)
Constant	−0.037
	(0.246)
Observations	4,296
R^2	0.138
Adjusted R^2	0.136
Residual Std. Error	0.878 (df = 4287)
F Statistic	85.434*** (df = 8; 4287)

Note: *p<0.1; **p<0.05; ***p<0.01

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