Effects of rest and tiredness on soccer and tennis match outcomes

Chloé Lepert

8/9/2018

1 Introduction

Many factors influence athletic performance in competitions. Most models predicting results focus on the strength of the team and the location of the competition, but other factors that players and teams can control could have an impact. In this paper, we look at how rest and tiredness impact 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 game load of a team in the past month, two months, and three months impact the performance of a team?

For tennis, we look at the effect of the previous match attributes on match outcomes. We ask the following questions:

1. How does the length of the previous match impact performance?

2. How does this effect vary by game attributes?

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

1.2 Football and Tennis

Before explaining our motivation, we will first give an overview of two sports so that the reader may better understand our motivation.

1.2.1   English Football

English soccer is comprised of multiple competitions.

- The top 20 teams in England play in the **Premier League**. They play each other twice per season: once at home and once away resulting in 380 games played. Each match earns teams points; a win gets a team 3 points, a draw 1 point, and loss 0 point. The team with the most points win. The bottom two teams are relegated to the EFL Championship. The top two teams from EFL Championship advance to the Premier League. The third to last team in the Premier League and the third team in the EFL Championship playoff to play in the Premier League. [**?**]

- The **FA cup** is a knockout tournament open to teams in 10 levels of English football. Teams enter the tournament at different times depending on the level they play in. Teams that play in the Premier league enter the 3rd round in January with the hope of making it to the final played in May. [**?**]

- The **EFL cup** is a tournament similar to the FA cup but open only to levels 1 through 4. Premier League teams enter in August. Matches continue until the final in February. Teams playing in European competitions do not play in this cup [**?**]

There are two European competitions played with teams from different European leagues:

- The top teams play in the **Champions League**. It starts with a group stage of 32 teams in which mini round robins are run in each group to determine which 16 teams go on to the tournament. Teams not advancing to the tournament are transferred to the Europa league. In the tournament stage, teams meet twice: once home and once away. Since only one team can move forward to the next round, the return leg can go overtime to decide who win. The final is played in one match in a location determined ahead of time. [**?**] In England the top three premier league teams automatically qualify for the Champions League. The fourth-place team qualifies for a playoff round. [**?**]

- The **Europa league** is similar to the Champions league in format but played with the top teams not qualifying for the champions league. Depending on how a team qualifies it enters the competition at different rounds. [**?**] 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. [**?**]

Depending on the year, the schedule is inter-spaced with 2-week **international breaks**, in which players are called to represent their national team in international friendlies or qualifying matches for continent or world cups.

The wide variety of competition that a team can play in means that teams have different schedule loads as shown in the following figure.

Distribution of the number of games played by a teams over one season
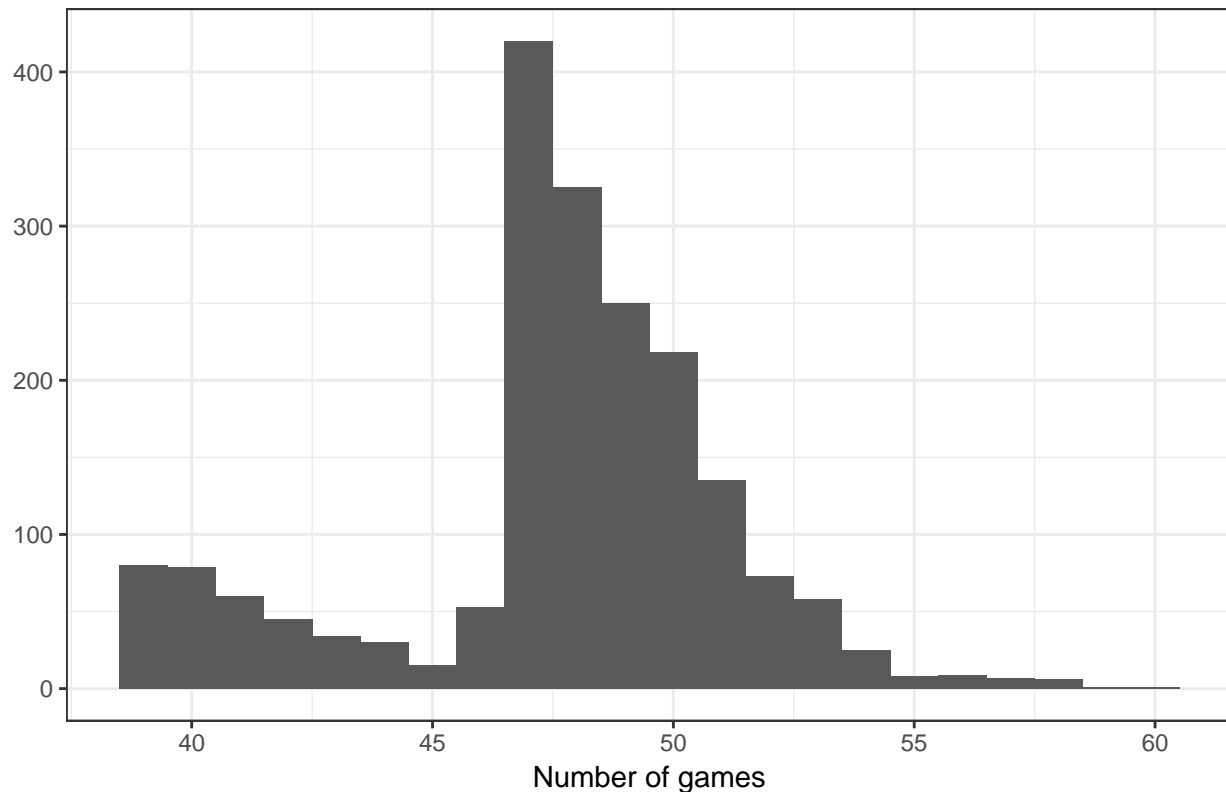
1.2.2   Men's professional tennis

Throughout the season, players play in a variety of tournaments. Performance in tournaments determines the number of points earned towards the players' world rankings. Tournaments have different importances influencing how many points they grant players.

A player's ranking determines a player's seeding in individual tournaments. Officials design seedings to have top players face each other as late in the tournament as possible.

Professional tennis is played in matches of 3 or 5 sets. The four Grand Slam tournaments: Australian Open, Rolland Garros, Wimbledon, and US open are played in 5 sets, while the rest of tournaments are played in 3 sets. The first player to win 2 or 3 sets wins the match and moves on to the next round.

The winning mechanism allows for games of different time length; the distribution of which is shown below.

Distribution of tennis match length
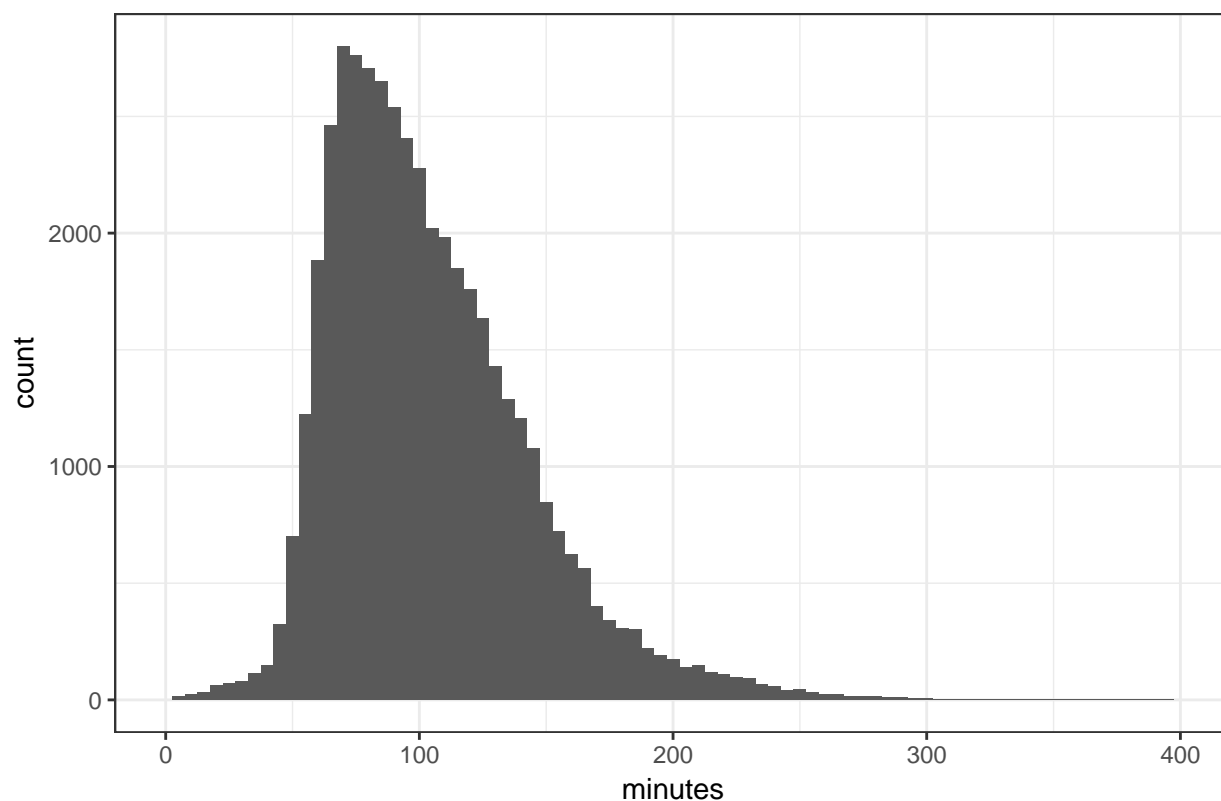
A set is won by winning at least six games and having a two-game lead resulting in sets of different intensities; a set could be won 6-1, of 10-8. The distribution of the number of games is shown below along.

Distribution of the number of games in tennis matches by number of set
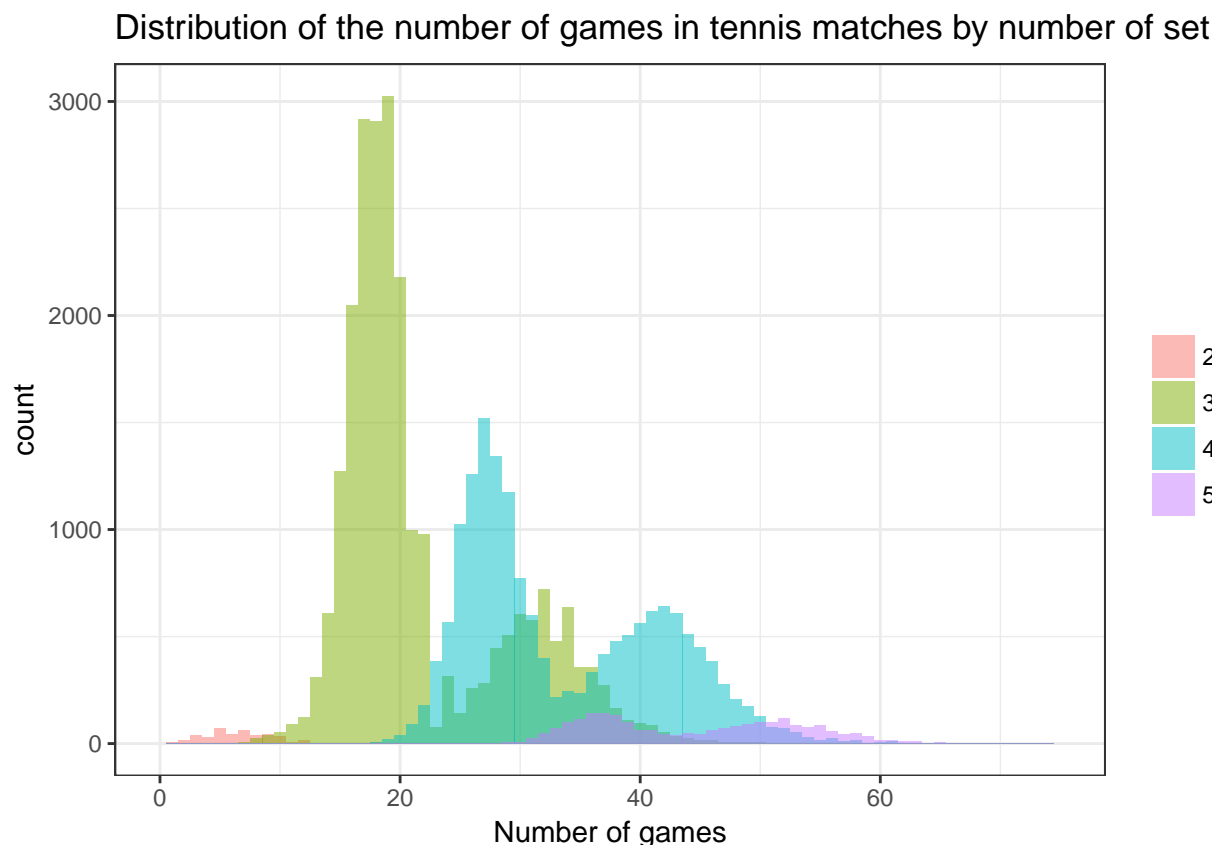
1.3 Relevance

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 by allowing teams that have the break to be more refreshed. [**?**]

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 by letting them schedule their games earlier in the weekend.

In tennis, seedings are designed so that players' pathes to the finals are as fair as possible. The tournament is set up at the beginning of the tournament such that assuming the highest ranked player always wins, in each round, the highest ranked player always plays the lowest ranked player. Sometimes a low ranked player will upset high ranked player which can cause an easier path for future players.

Here we look at whether the length of the previous match on a player's path impacts performance on the current game. While some of the influence of a match's length are outside a player's control, for example how hard their oponent is to beat, other attributes might not

be. An effect of previous matches on present matches could affect how matches are played. If player's know that long matches hurt them in the long run, they might work harder to win quickly. Of course, we also need to measure the LR effect of working harder before making such a suggestion.

1.4   Approach

While we would ideally want to look at how rest affects performance in European competitions doing so is difficult because the teams that play each other in European competitions do not do so frequently enough for us to calculate and control for team ability on an even scale. We will instead model the outcome of Premier League matches, which are set up such that we can easily control for team ability. We will evaluate how matches in competitions other than the premier league influence premier league matches.

This approach assumes that European competition games are similar to 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.

For tennis, we measure the effects of previous match attributes on the next match. We look at how these effects differ by surface and whether the match is best of 3 or best of 5.

1.5   Literature review

1.5.1   Rest time

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

Over the past decades, Carlos Lago Penas conducted a series of studies 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 rate for football players who played two matches a week compared to players playing one match a week.

1.5.2   Modeling football match outcomes

The sports statistic community quickly converged on the best way to model soccer scores. In 1982, M. J. Maher introduced two independent Poisson distributions as a way to model soccer scores. He proposed using a team's attacking strength and its opponent's defensive weakness

as predictors of the number of goals scored. Maher also found that using a Bivariate Poisson distribution with a correlation of 0.2 improved his model's fit. [**?**]

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

We will also look at the proportional odds ordered logistic model which predicts the probability of a win, tie, or loss.

1.5.3   Tennis match length

In 2007, Pereir, Nakamura, and de Jesus tracked physical performance -distance covered in a set and stroke proficiency in the first two sets of a match and found no decrease in physical performance.

In 2015, Goosens, Kemperas, and Kosing looked at how the difference in number of sets in the previous match impacts outcome in the current match in Grand Slams. They found a significant decrease in probability of winning for a difference in sets number of 1 for women and 2 for men.

1.5.4   Predicting tennis matches

Most models aimed at tennis betting use the hierarchical structure of tennis matches set-game-point to build stochastic models. In 2015, Sipky proposed using historical data about the players and a neural network and achieved a 75% performance improvement over existing models. In 2003, Klassen and Magnus proposed a logistic model to forcast tennis matches. We chose this model in our analysis as it easily lends itself to inference.

1.6   Data

1.6.1   Tennis

1.6.1.1   Data set

The data was obtained from Kaggle and contains outcomes and attributes of ATP (Association of Tennis Professionals) Mens tennis matches between 2000 and 2017. We split each game into two observations, one for each player.

1.6.1.2   Response variable

The response variable is whether or not a player won a game. As players have to either win or lose, the response variable is distributed 50-50 win-lose.

1.6.1.3 Predictors

The predictor is the length of the previous match played by the player in hours Three-quarters of the games are less than 2 hours long, but games can go on for many more hours. About five percent of games go on for more than 3 hours.

1.6.1.4 Control variables

We control for player ability by looking at the player's rank at the beginning of the tournament as well as the number of points earned in order to achieve the ranking. Both controls are important in predicting winners. They differ slightly in that the player ranked number 1 could be leading by a few points or hundreds of point. The broader scale of ranking points provides a better proxy for underlying player ability.

We also control for whether or not a player is seeded. Seeded players tend to be the top 32 players in the tournament and do not have to play as many early rounds.

We also look at whether or not the effects differ by surface or by the number of sets needed to win.

1.6.1.4.1 Surface

Tennis is either played on Carpet, Clay, Grass, or Hard. Players will have preferences for different surfaces, and some surfaces are often cited as causes for injuries (cite). The distribution of surfaces is shown below.

Table 1: Distribution of matches by surface

	Var1	Freq
1	Carpet	$1,489$
2	Clay	$15,456$
3	Grass	$4,466$
4	Hard	$23,876$

1.6.1.4.2 Sets needed to win

Most men's tennis matches are won by a player winning two sets. The four major tournaments: Rolland Garros, Wimbledon, US Open, and Australian open require three sets to be won in order to win

1.6.1.5 Games looked at

We look at all matches for which both players played their previous match in the same tournament. This allows us to make sure that the previous match happened within a

reasonable number of days and that surface and best of quantity are the same between player histories. We remove any match including and unranked player

1.6.2 Soccer

1.6.2.1 Data set

The R package "engsoccerdata" provides us with scores from matches in England as well as European competition matches. [**?**] We limit ourselves to matches occurring in seasons 1995 through 2015. The design of English Football stays consistent throughout this time range, and the data set is complete for these years.

1.6.2.2 Response variable

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

Distribution of goals overlayed with expected number of goals assuming a poisson distribution

Home goals

Mean: 1.51 Variance: 1.71

Visitor goals

Mean: 1.11 Variance: 1.23
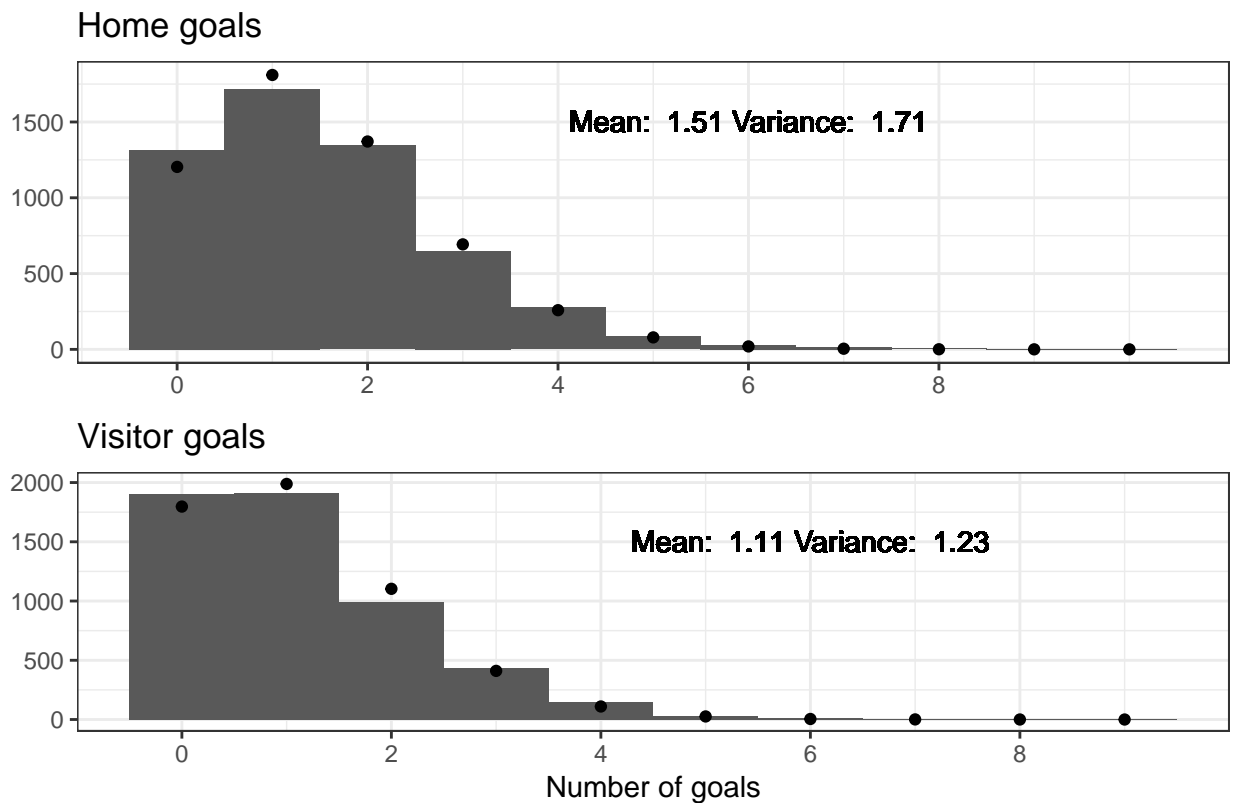
Number of goals

Most models for the number of goals by a team in a game assume the number of home and away goals follow a Poisson distribution. We see that this approximately true. There tend to be 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 2: Distribution of game outcomes

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

1.6.2.3 Predictors

Using this dataset, we calculate the number of days since the previous game. Before performing our analysis on the effect of rest time, we remove games in which either of the team's previous game was over eight days ago. If the team's last non-international game was more than eight days prior, it is possible that an international break occured which we have no way of controlling for.

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

Table 3: Distribution of rest time (in days) for home and away teams

		Visitor						
		2	3	4	5	6	7	8
	2	128	9	1	0	3	0	1
	3	12	343	155	22	38	121	7
	4	0	165	238	43	38	91	61
Home	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 whether or not a team is "more" (6-8 days) or "less" (1-5 days) rested and found the following distribution of rest.

1.6.2.4 Control variables

Using Premier Leauge games, we calculate each team's attacking and defensive weakness. [?] The attacking strength is the ratio of the average number of goals scored 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

Table 4: Distribution of rest time for home and away teams

| | | Away | |
		1-5 days	6-8 days
Away	1-5 days	$1,256$	512
	6-8 days	479	$2,049$

average number of goals conceded in the league that season. It is generally believed that the performance of a team differs depending on whether the team is home or visiting; therefore we calculate Attacking Strength and Defensive Weakness 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, since they played against a different set of teams. The distribution of attacking and defensive strengths is shown below.

Distribution of attacking strength and defensive weakness for home and away
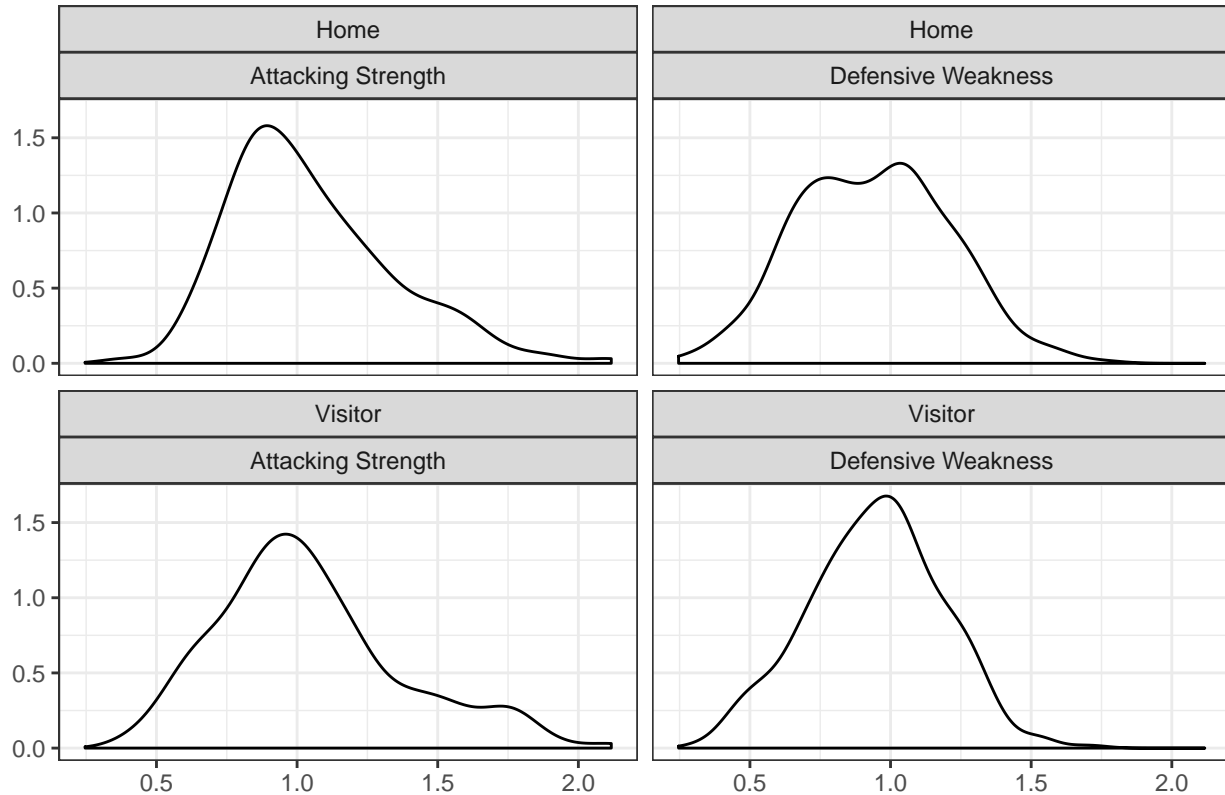
We also control for the number of games a team plays each season. Better teams will last longer in playoff competitions such as the FA Cup, Europa League, and Champions League, that happen concurrently with premier league games. These teams will play more games per season and thus on average have lower rest times.

1.6.2.5 Matches looked at

As described in the previous subsections there are two points at which we remove premier league games from our data set.

1. Matches with recently promoted teams. Since we use the previous premier league season to calculate the team's attacking strength and defensive weakness we cannot calculate these quantities for recently promote teams. Removing these matches removes 2540 out of 7980 matches (32%).

2. Matches with over eight days of rest. It is likely that an international match happened during the break which we do not have in our data. Leaving these matches in would lead to low rest matches presenting themselves as high rest matches. Removing these matches removes 1144 out of 5440 matches (21%).

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 its 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
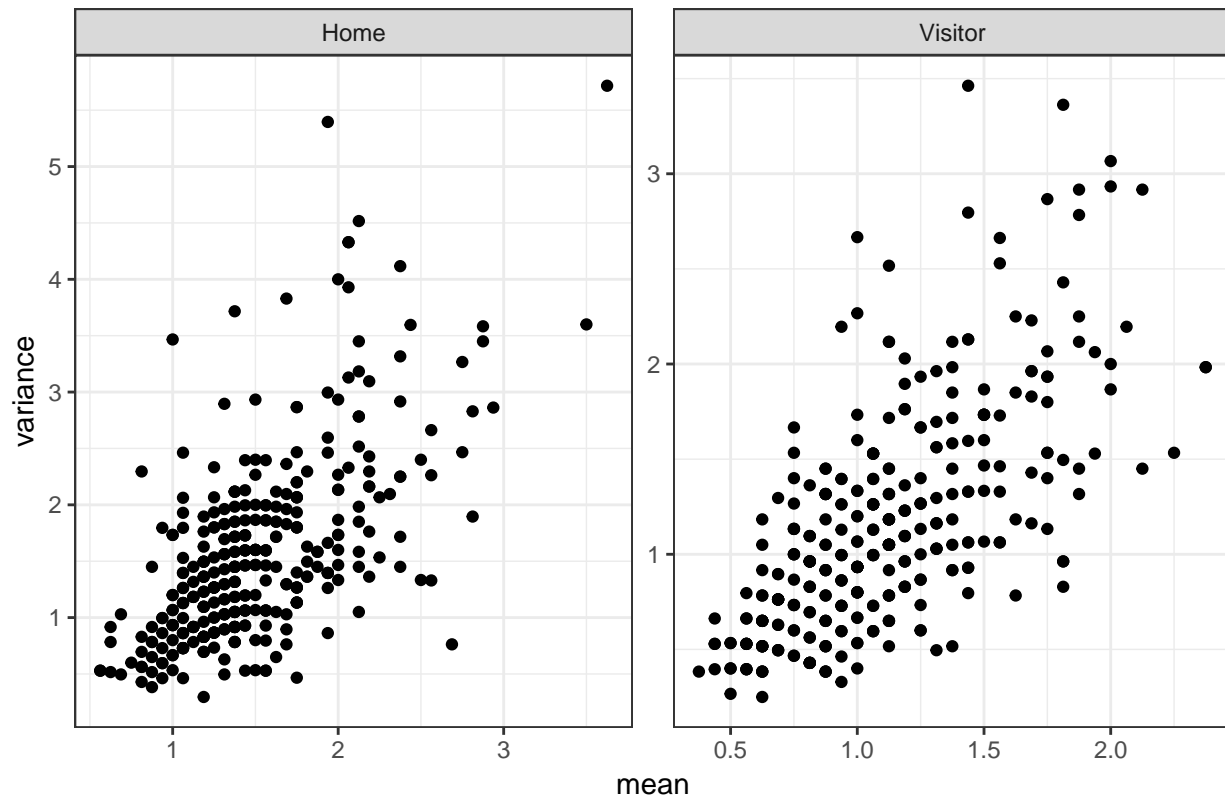
2.2 Linear model for goal difference

The square root of a Poisson distribution can be approximated by a normal 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 square root 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
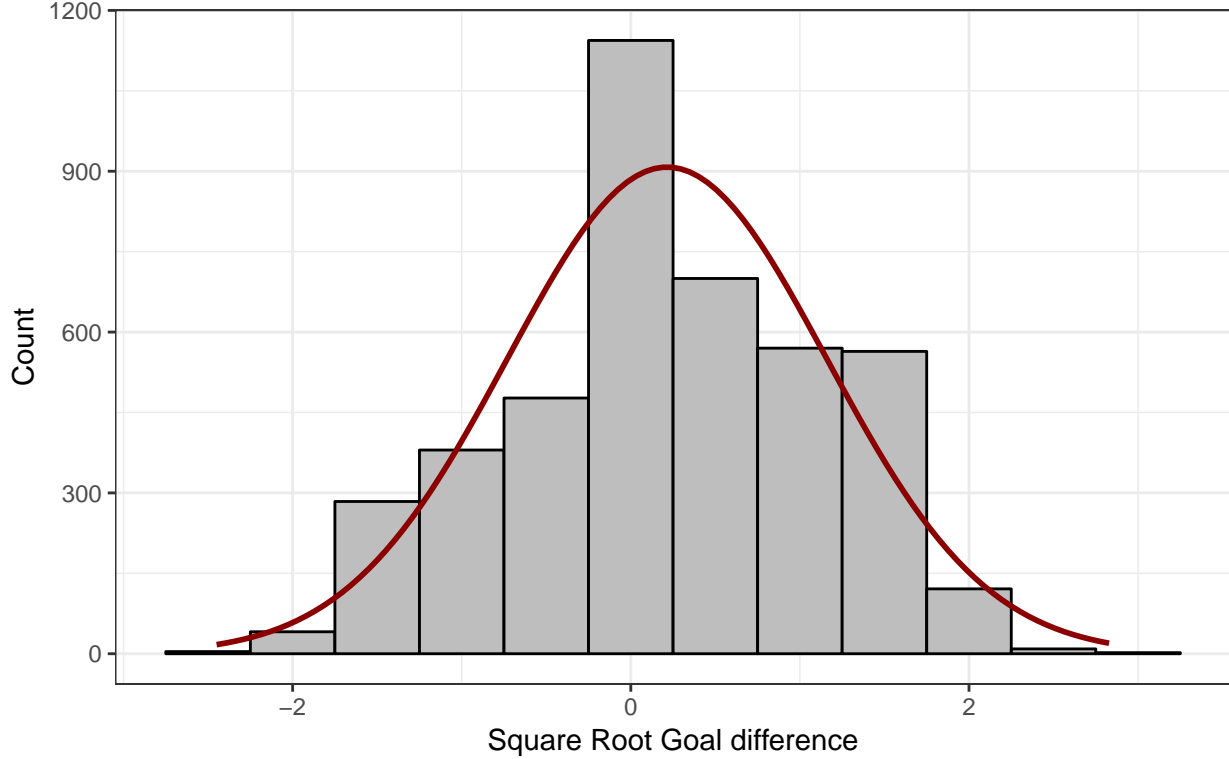
Our model aims to find a linear combination of our predictors that best approximates the square root goal difference.

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

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

2.3 Proportional odds cummulative logit model

We have three possible outcomes: a home win, tie, or visitor win. Each outcome has a probability π_i of happening. The probabilities of the three outcomes sum to 1 as no other outcome is possible. $\pi_h + \pi_t + \pi_v = 1$

The probability that home loses is $1 - \pi_h - \pi_t = \pi_v$ and it's log odds are $L_v = \log\left(\frac{\pi_v}{\pi_h + \pi_t}\right)$

The probability that home loses or ties is $1 - \pi_h = \pi_v + \pi_t$ and it's log odds are $L_t = \log\left(\frac{\pi_v + \pi_t}{\pi_h}\right)$

The log odds of these two events are assumed to be a linear combination of the predictors; $L_v = \alpha_v + X\beta_v$ and $L_t = \alpha_t + X\beta_t$.

14

In the proportional odds model we require the coefficients β_v and β_t to be the same; $L_v = \alpha_v + X\beta$ and $L_t = \alpha_t + X\beta$.

The underlying assumption is that there is an unobserved variable which determines match outcomes. Here this variable is the ability of a home team to win against a given opponent. The predictors X change our belief as to how this ability is distributed.

Distribution of ability to win a match depending on whether or not a feature
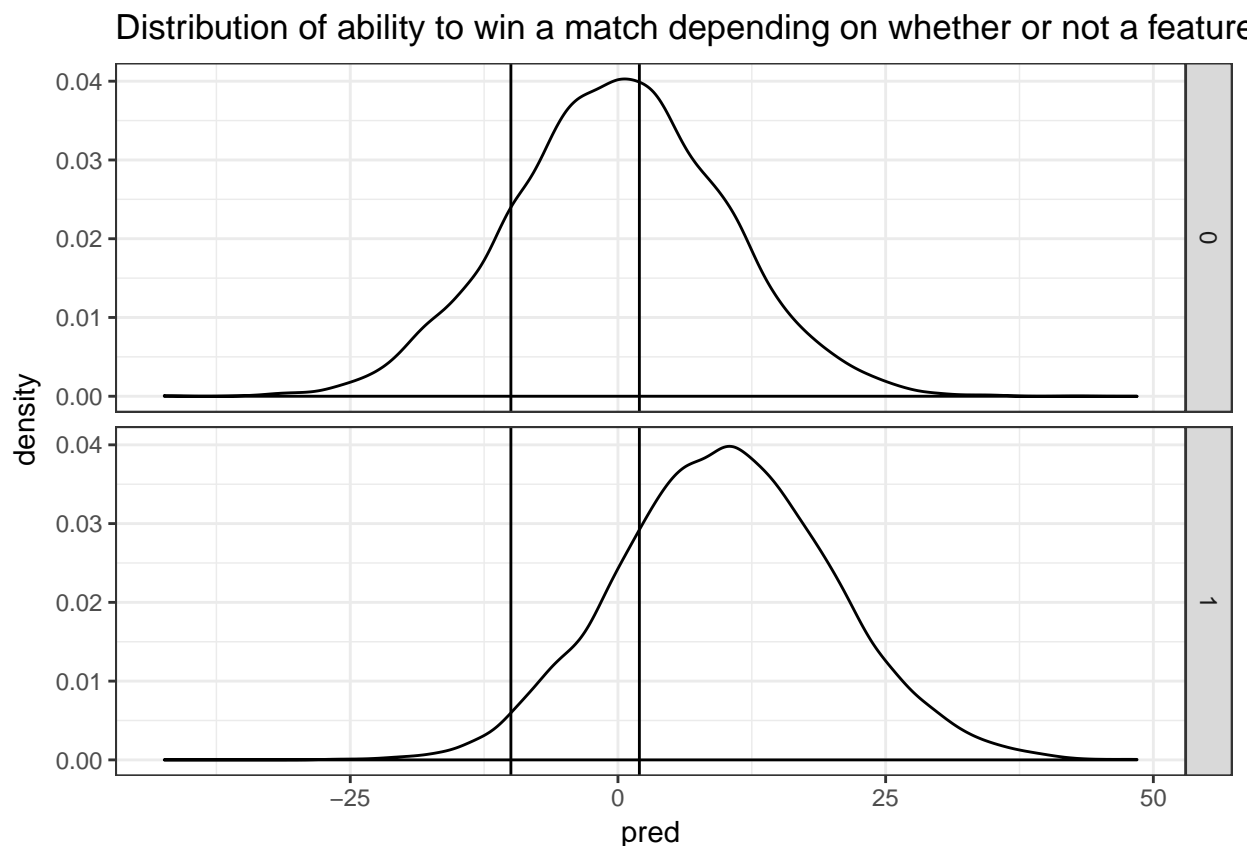
α_v can be thought of as the threshold in ability to go from a tie or home win to a visitor win and α_t as the threshold to go from a home win to a tie or visitor win or tie. Depending on the distribution of underlying ability, the probability of each outcome changes. In the figure above if the feature is matched, the probability of a win is 78% but if the feature is not matched the probability of a win is 42%.

The two outcome case of this model is known as logistic regression. We will use it to model the outcome of tennis matches.

2.4   Logistic regression

Logistic regression allows us to model binary outcomes where one outcome, here a win, happens with probability p and the other, here a loss, happens with probability 1-p. Logistic regression assumes that the log odds of the probability of the event are given by a linear combination of the predictors. The model can be written as $\log\left(\frac{p}{1-p}\right) = X\beta$ where the first column of X is filled with ones for the intercept, and the other columns are the predictors.

3 Results

3.1 Simple Poisson GLM

Table **??** summarizes the GLM models for the number of home and away goals. Rest comes in the model either as the number of days since the previous game (1) and (3) or as whether or not the team had more than six days of rest (2) and (4). In neither of the four models do we see rest as affecting 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 too 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 5: 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		0.047		−0.044
		(0.031)		(0.036)
Opp_rest_bin		−0.052*		0.043
		(0.031)		(0.037)
Team_att_str	0.371***	0.377***	0.427***	0.425***
	(0.050)	(0.050)	(0.053)	(0.053)
Opp_def_weak	0.274***	0.280***	0.238***	0.239***
	(0.061)	(0.061)	(0.061)	(0.061)
Team_load	0.015***	0.017***	0.011***	0.010***
	(0.003)	(0.003)	(0.003)	(0.003)
Opp_load	−0.020***	−0.021***	−0.024***	−0.023***
	(0.003)	(0.003)	(0.004)	(0.004)
Constant	−0.036	−0.077	−0.025	−0.041
	(0.205)	(0.199)	(0.231)	(0.223)
Observations	4,296	4,296	4,296	4,296
Log Likelihood	−6,597.132	−6,595.641	−5,804.871	−5,804.193
Akaike Inf. Crit.	13,208.260	13,205.280	11,623.740	11,622.390

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

3.2 Linear model for goal difference

Table 6 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 6: 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

3.3   Proportional odds cummulative logit model

The odds of moving from a visitor win to a tie or a tie to a home win increase when the home team's attacking strength, home team's game load, or visitor's team defensive weakness increases. The opposite happens when the teams are reversed. The directionality of these findings agrees with the other models. The proportional odds cumulative logit model does not find any effect of rest, whether measured in days or as a binary variable.

Table 7: Ordered logistic model

	Dependent variable:	
	outcome	
	(1)	(2)
Team_rest	−0.016	
	(0.022)	
Opp_rest	0.006	
	(0.022)	
Team_rest_bin		−0.002
		(0.074)
Opp_rest_bin		−0.039
		(0.074)
Team_att_str	0.864***	0.870***
	(0.130)	(0.130)
Opp_def_weak	0.700***	0.707***
	(0.145)	(0.145)
Opp_att_str	−0.717***	−0.722***
	(0.111)	(0.111)
Team_def_weak	−0.490***	−0.490***
	(0.125)	(0.125)
Team_load	0.042***	0.043***
	(0.008)	(0.008)
Opp_load	−0.034***	−0.035***
	(0.008)	(0.008)
Observations	4,296	4,296

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

3.4 Logistic GLM

In the simplest model looking at how the difference in minutes of the previous matches affects the probability of winning we see that having played a longer previous match than one's opponent increases the probability of winning. We then look for the diminishing effect of the number of prior minutes played by looking at the importance of minutes played squared and found no diminishing effects.

When we control for the difference in ranking points between the two players the effect disappears. Further, controlling for whether or no a player or his opponent was seeded further helps predict tennis match outcomes.

Table 8: Logistic model for the probability of winning a tennis match

	Dependent variable:			
	won			
	(1)	(2)	(3)	(4)
delta_hours	0.027***		0.010	−0.003
	(0.008)		(0.008)	(0.008)
poly(delta_hours, 2)1		6.991***		
		(2.001)		
poly(delta_hours, 2)2		0.000		
		(2.038)		
delta_rank_pts			0.494***	0.351***
			(0.006)	(0.007)
player_seed				0.447***
				(0.017)
opp_seed				−0.447***
				(0.017)
Constant	0.000	0.000	−0.000	0.000
	(0.007)	(0.007)	(0.007)	(0.010)
Observations	90,574	90,574	90,574	90,574
Log Likelihood	−62,775.010	−62,775.010	−57,235.310	−56,624.180
Akaike Inf. Crit.	125,554.000	125,556.000	114,476.600	113,258.400

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

When we look at the effect by best of quantity, we find a positive effect on winning probability for best of three matches and a negative effect on winning probability for best of five matches.

We then look at how surface plays a role we find that prior match length increases the probability of winning for clay surfaces and decreases the probability of winning for grass surfaces.

We then split the surface effects by best of quantity and see that clay only increases the probability of winning on best of three matches and grass only decreases the probability of winning for best of 5 matches. As previously seen the effects are positive for best of three matches and negative for best of five matches.

Table 9: Effect of previous match length on winning probability by match type

	Dependent variable:		
	won		
	(1)	(2)	(3)
delta_rank_pts	0.350***	0.351***	0.350***
	(0.007)	(0.007)	(0.007)
delta_hours:surfaceCarpet		−0.029	
		(0.047)	
delta_hours:surfaceClay		0.033**	
		(0.015)	
delta_hours:surfaceGrass		−0.049**	
		(0.023)	
delta_hours:surfaceHard		−0.011	
		(0.012)	
player_seed	0.446***	0.447***	0.446***
	(0.017)	(0.017)	(0.017)
opp_seed	−0.446***	−0.447***	−0.446***
	(0.017)	(0.017)	(0.017)
delta_hours:as.factor(best_of)3	0.024**		
	(0.009)		
delta_hours:as.factor(best_of)5	−0.096***		
	(0.018)		
delta_hours:as.factor(best_of)3:surfaceCarpet			−0.028
			(0.048)
delta_hours:as.factor(best_of)5:surfaceCarpet			−0.156
			(0.631)
delta_hours:as.factor(best_of)3:surfaceClay			0.048***
			(0.016)
delta_hours:as.factor(best_of)5:surfaceClay			−0.043
			(0.036)
delta_hours:as.factor(best_of)3:surfaceGrass			0.055*
			(0.030)
delta_hours:as.factor(best_of)5:surfaceGrass			−0.183***
			(0.035)
delta_hours:as.factor(best_of)3:surfaceHard			0.006
			(0.013)
delta_hours:as.factor(best_of)5:surfaceHard			−0.075***
			(0.025)
Constant	0.000	−0.000	0.000
	(0.010)	(0.010)	(0.010)
Observations	90,574	90,574	90,574
Log Likelihood	−56,606.370	−56,618.790	−56,598.650
Akaike Inf. Crit.	113,224.700	113,253.600	113,221.300

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

4 Discussion

4.1 Soccer

4.1.1 Comparison of models

- Team effects agree
- Which model is most valid

4.1.2 Rest as days or as binary

4.1.3 Importance of game load

4.2 Tennis

4.2.1 Importance of controlling for ranking points

In the simplest model for the probability of winning a tennis match the difference in the length of the player's previous matches are highly significant. However, when we add the difference in the players' ranking points and binary variables for whether or not the players' involved are seeded, the significance of the effect disappears suggesting some correlation between the predictors

Indeed we find a correlation of 0.070 with a p-value less than 1 e-15 between a player's ranking points and the time needed to win a match. We also find that seeded players enter matches having played on average 3.87 minutes (SE 0.28) more in their previous match than their unseeded opponents. Both of these findings suggest that better players play longer matches.

One hypothesis for why better players play longer match is that they are more able to put up a fight and last longer against strong opponents. If that is true, we expect the length of matches of better players to increase more with the strength of their opponents than weaker players.

In the Poisson model below in which we assume that the length of a tennis match is Poisson distributed according to a parameter which is linear in the predictors, we see that the length of seeded players matches increases with the ranking of their opponents more than the length of the matches of unseeded players. This supports our hypothesis that better players (proxied by seeded players), put up stronger fights against better opponents than unseeded players do.

Table 10: Effect of opponent's ranking points on match length by player seed

	Dependent variable:
	minutes
player_seedFALSE:opp_rank_points	0.005***
	(0.0002)
player_seedTRUE:opp_rank_points	0.025***
	(0.0002)
Constant	4.629***
	(0.0004)
Observations	90,574
Log Likelihood	−924,754.200
Akaike Inf. Crit.	1,849,514.000

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

Since a player's ability is correlated with the length of matches that player plays, it is important to control for it so that we can truly extract the effect of the length of the previous match on match outcomes. We used both rank points and seeded status of the players as controls.

4.2.2 How might the effect not be valid?

4.3 Comparison of sports

4.3.1 Why is it easier to pick up effects on tennis

- One player vs. many. Football manager can control how tired his players are by rotating players something which tennis players cannot do.

- Larger variation in the predictor for tennis

5 Conclusion

6 Appendix

7 References

8 Notes

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