# Measuring the effect of rest and tiredness in soccer and tennis outcomes

Chloé Lepert
Masters Paper Presentation
1 November 2017

# Agenda

- Soccer

  - Question asked

  - Literature review

  - Data

  - Models and their outcomes

  - Soccer conclusions

- Time permitting: Tennis

# How does the number of days since a previous match impact a soccer team's performance?

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 |

# Existing research on tiredness in soccer focuses on ability to complete physical tasks

- Carlos Lago Penas looked at distance run at certain speed by players over a succession of matches with low rest between matches and found no differences

- Dupond et. al. found that soccer players playing 2 times a week had a higher rate of injury but did not have worse physical performance

# Data

- R package "engsoccerdata" provides dates, and results for all matches involving English team in the past century

  - We look at matches between 1995 and 2015 to ensure data completeness

  - Variables we look at include: the date, home team, visitor team, number of home goals, number of visitor goals

# Data - Rest days - 1

- The first division of English soccer team plays in the Premier League along with other English championship (FA Cup,…) and European championships (Europa League Champions League)

  - In the Premier league each team plays each other twice in a round robin fashion

  - All other games are part of tournaments in which the best team moves forward

- We use all matches to compute the rest days of teams

- Calculate the effect of number of rest days on Premier league matches

# Data - Rest days

| | Sun | M | Tue. | W | Thu. | F | Sat. | Sun. | M | Tue. | W | Thu. | F | Sat. | Sun. |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Man City | PL - 11 | | | | | | PL -12 | | | | | | | PL - 13 | |
| Arsenal | PL - 11 | | | CL - 1/4 | | | PL - 12 | | | | | | | | PL - 13 |
| Man United | PL - 11 | | | | | | | PL - 12 | | | | | | | PL - 13 |

- 2 variables for rest:

    - Number of days of rest

    - Whether or not the number of days of rest exceeds 5 (high rest) or not (low rest)

- On day 12, Man City plays Arsenal at home. Man City has 6 days of rest (high rest) and Arsenal has 3 (low rest)

- On day 13, Arsenal plays Man Unites at home. Man United has 7 days of rest (high rest) and Arsenal as 8 days (high rest)

# Data - Control variables

- Total number of matches played in a season

- Attacking strength and defensive weakness

    - Attacking strength = average number of goals scored by a team / average number of goals scored in the premier league

    - Defensive weakness = average number of goals conceded by a team / average number of goals conceded in the premier league

    - Calculate attacking strength and defensive weakness using the previous year's performance

# Data - Response

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



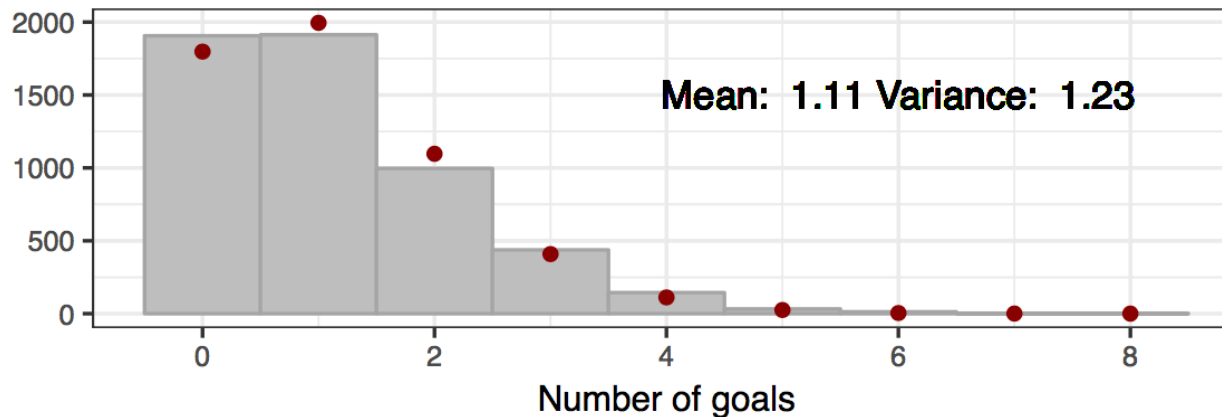Mean: 1.51 Variance: 1.71

Visitor goals



Mean: 1.11 Variance: 1.23
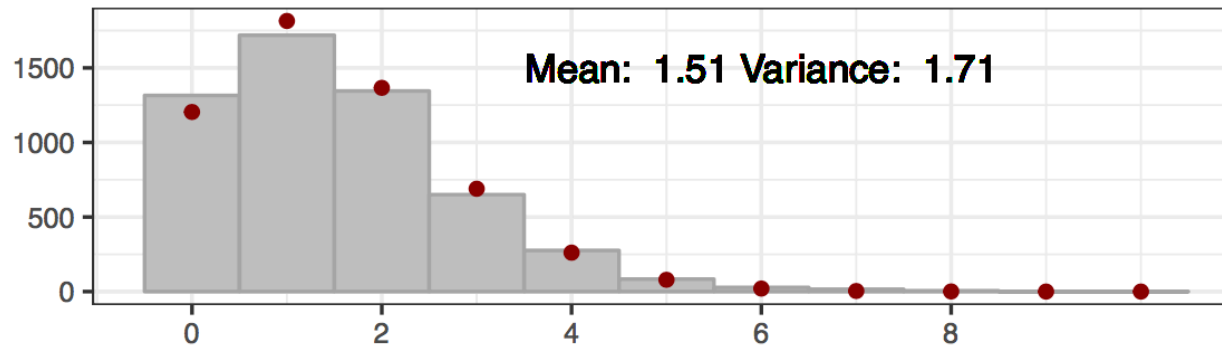
Number of goals

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 |

# Models

- GLM with Poisson Link

- Linear model for goal difference

- Proportional odds cumulative logit model

- Bivariate Poisson

# GLM with Poisson link

- 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 $\lambda_j$ is a linear combination of the predictors $X_j$: $\lambda_j = X_j \beta_j$

## Table 5: Generalized Linear Models with Poisson link

| | Dependent variable: | | | |
| --- | --- | --- | --- | --- |
| | Home goals | | Visitor goals | |
| | (1) | (2) | (3) | (4) |
| Team rest (days) | −0.004 (0.009) | | −0.007 (0.011) | |
| Opponent rest (days) | −0.001 (0.009) | | 0.005 (0.011) | |
| Team rest > 5 days | | 0.047 (0.031) | | −0.044 (0.036) |
| Opponent rest > 5 days | | −0.052* (0.031) | | 0.043 (0.037) |
| Team attacking strength | 0.371*** (0.050) | 0.377*** (0.050) | 0.427*** (0.053) | 0.425*** (0.053) |
| Opp. defensive weakness | 0.274*** (0.061) | 0.280*** (0.061) | 0.238*** (0.061) | 0.239*** (0.061) |
| Team load | 0.015*** (0.003) | 0.017*** (0.003) | 0.011*** (0.003) | 0.010*** (0.003) |
| Opponent load | −0.020*** (0.003) | −0.021*** (0.003) | −0.024*** (0.004) | −0.023*** (0.004) |
| Constant | −0.036 (0.205) | −0.077 (0.199) | −0.025 (0.231) | −0.041 (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

# Bivariate Poisson

- Assume that $G_j \sim Poisson(\lambda_j + \lambda_g)$.

- The parameter $\lambda_g$ is a linear combination of the predictors $X$: $\lambda_g = X_g \beta_j$

The probability distributions for the number of goals by the home and away team is given by:

$$P_{BP}(G_h = g_h, G_a = g_a | \lambda_h, \lambda_a, \lambda_g) = e^{-(\lambda_h + \lambda_a + \lambda_g)} \frac{\lambda_h^{g_h}}{g_h!} \frac{\lambda_a^{g_a}}{g_a!} \sum_{i=0}^{min(g_h, g_a)} \binom{g_h}{i} \binom{g_a}{i} i! \left(\frac{\lambda_g}{\lambda_g \lambda_a}\right) \quad (1)$$

Table 8: Bivariate model for number of goals scored by each team

| | | (1) | (2) |
|---|---|---|---|
| Home | Intercept | -0.254***(0.017) | -0.252***(0.018) |
| | Visitor defensive weakness | 0.427***(0.005) | 0.418***(0.005) |
| | Visitor load | -0.021***(0) | -0.019***(0) |
| | Visitor rest (days) | | -0.004***(0.001) |
| | Visitor rest >5 days | -0.081***(0.003) | |
| | Home attacking strength | 0.451***(0.004) | 0.45***(0.004) |
| | Home load | 0.014***(0) | 0.013***(0) |
| | Home rest (days) | | 0.001 (0.001) |
| | Home rest >5 days | 0.06***(0.002) | |
| Visitor | Intercept | 0.1***(0.02) | 0.046**(0.022) |
| | Visitor Attacking strength | 0.5***(0.004) | 0.506***(0.004) |
| | Visitor load | 0.008***(0) | 0.008***(0) |
| | Visitor rest (days) | | -0.01***(0.001) |
| | Visitor rest >5 days | -0.075***(0.003) | |
| | Home defensive weakness | 0.275***(0.005) | 0.278***(0.005) |
| | Home load | -0.028***(0) | -0.028***(0) |
| | home rest (days) | | 0.01***(0.001) |
| | Home rest >5 days | 0.055***(0.003) | |
| Game | Intercept | 0.156 (0.179) | 0.872***(0.164) |
| | Visitor attacking strength | -0.891***(0.054) | -0.891***(0.059) |
| | Visitor defensive strength | -2.785***(0.073) | -2.641***(0.083) |
| | Visitor load | -0.008**(0.003) | -0.012***(0.004) |
| | Visitor rest | | 0.054***(0.011) |
| | Visitor rest >5 days | 0.456***(0.086) | |
| | Home attacking strength | -1.241***(0.08) | -1.289***(0.077) |
| | Home defensive weakness | -0.478***(0.054) | -0.505***(0.058) |
| | Home load | 0.057***(0.003) | 0.051***(0.003) |
| | Home rest (days) | | -0.07***(0.01) |
| | Home rest >5 days | -0.129***(0.035) | |

# Linear Model for goal difference

- Square root of a poisson random variable can be approximated as a normal r.v.

- Difference of square root goals can be modeled by linear model

Table 6: Linear model for the difference in goals scored

|  | Dependent variable: |
| --- | --- |
|  | Goal difference |
| Team rest (days) | −0.005 (0.010) |
| Opponent rest (days) | −0.003 (0.010) |
| Team attacking strength | 0.399*** (0.057) |
| Team defensive weakness | −0.277*** (0.057) |
| Opp. attacking strength | −0.380*** (0.050) |
| Opp. defensive weakness | 0.364*** (0.066) |
| Team load | 0.019*** (0.003) |
| Opponent 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 |

# Proportional Odds Logit Cumulative model

We have three possible outcomes: a home win, tie, or visitor win. Each outcome has a probability $\pi_i$ 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 its 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 its 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 $L_t = \alpha_t + X\beta_t$.

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

# Proportional Odds Logit Cumulative model



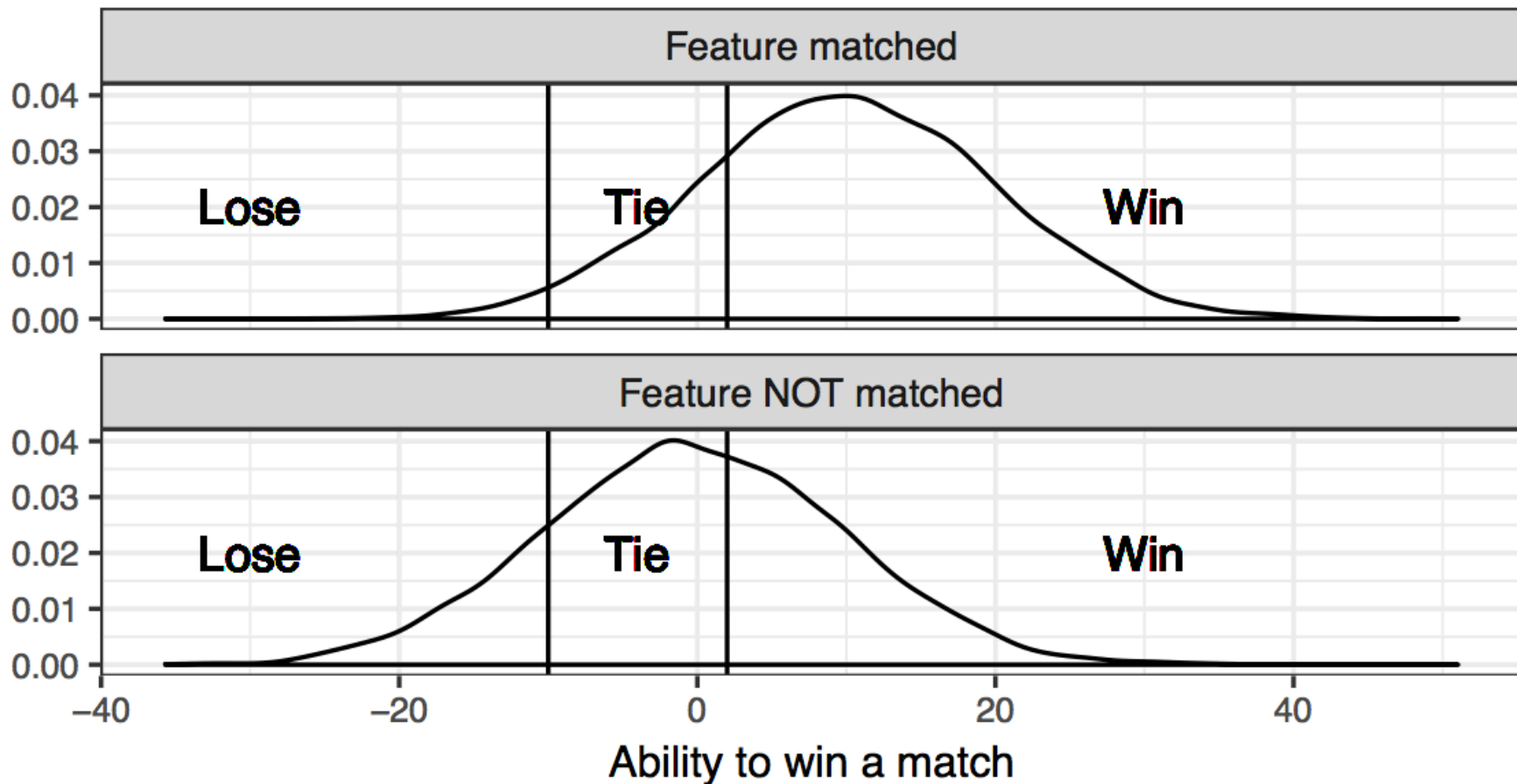Distribution of ability to win a match depending on whether or not a feature is matched

## Table 7: Ordered logistic model

| | Dependent variable: | |
| --- | --- | --- |
| | Probability of winning | |
| | (1) | (2) |
| Team rest (days) | −0.016 (0.022) | |
| Opponent rest (days) | 0.006 (0.022) | |
| Team rest > 5 days | | −0.002 (0.074) |
| Opponent rest > 5 days | | −0.039 (0.074) |
| Team attacking strength | 0.864*** (0.130) | 0.870*** (0.130) |
| Opp. defensive weakness | 0.700*** (0.145) | 0.707*** (0.145) |
| Opp. attacking strength | −0.717*** (0.111) | −0.722*** (0.111) |
| Team defensive weakness | −0.490*** (0.125) | −0.490*** (0.125) |
| Team load | 0.042*** (0.008) | 0.043*** (0.008) |
| Opponent load | −0.034*** (0.008) | −0.035*** (0.008) |
| Observations | 4,296 | 4,296 |

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

# Discussion

- Tried 4 types of model

  - Only one resulted in a significant effect for rest

  - Proportional odd logit and GLM with poisson link have advantage of easy explainability

- We do not find an effect of rest on match outcomes. Possible reasons why:

  - Soccer is a low scoring game, rest may have an impact on performance but not a big enough one to be measured in scores

  - Teams have 23+ players but at most 14 play in a single match. Managers may be controlling for rest in the teams they field.

# Tennis

Table 10: Effect of previous match length on winning probability by match type

|  | Dependent variable: | |
| --- | --- | --- |
|  | Win probability | |
|  | (1) | (2) |
| Diff. in ranking points | 0.351*** (0.007) | 0.352*** (0.007) |
| Player is seeded | 0.444*** (0.017) | 0.446*** (0.017) |
| Opponent is seeded | −0.444*** (0.017) | −0.446*** (0.017) |
| Surface: carpet - PMLD (hours) |  | −0.029 (0.047) |
| Surface: clay - PMLD (hours) |  | 0.033** (0.015) |
| Surface: grass - PMLD (hours) |  | −0.050** (0.023) |
| Surface: hard - PMLD (hours) |  | −0.011 (0.012) |
| Best of 3 - PMLD (hours) | 0.024** (0.009) |  |
| Best of 5 - PMLD (hours) | −0.096*** (0.018) |  |
| Constant | 0.000 (0.010) | −0.000 (0.010) |
| Observations | 90,620 | 90,620 |
| Log Likelihood | −56,641.040 | −56,653.550 |
| Akaike Inf. Crit. | 113,294.100 | 113,323.100 |

*Note:* $^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$

# Thanks

- Professor Stigler for advising me