

Modelling Football Results Using Match-specific Covariates

Gunther Schauberger* & Andreas Groll† & Gerhard Tutz*

We analyze the influence of various match-specific covariates on the results of football matches considering data from the German Bundesliga. Possible covariates are, for example, the ball possession or the running distance of both teams. For this purpose, the ordinal Bradley-Terry model is extended as follows:

$$\begin{aligned} P(Y_{i(rs)} \leq k | \mathbf{x}_{ir}, \mathbf{x}_{is}, \mathbf{z}_{ir}, \mathbf{z}_{is}) &= \frac{\exp(\theta_k + \gamma_{ir} - \gamma_{is})}{1 + \exp(\theta_k + \gamma_{ir} - \gamma_{is})}, \\ \gamma_{ir} &= \alpha_{r0} + \mathbf{x}_{ir}^T \boldsymbol{\beta} + \mathbf{z}_{ir}^T \boldsymbol{\alpha}_r. \end{aligned}$$

The response $Y_{i(rs)}$ represents an ordinal version of the outcome of match i between the teams r and s in $K = 5$ ordinal categories ($r \gg s$, $r > s$, $r = s$, $r < s$, $r \ll s$). The model can be embedded into the framework of cumulative logit models where θ_k represent the threshold parameters between the response categories $k = 1, \dots, K$. Instead of global strengths of the single teams γ_r , match-specific strengths γ_{ir} are considered depending on match-specific covariates \mathbf{x}_{ir} or \mathbf{z}_{ir} of team r in match i .

The covariates could be modeled using either global parameters $\boldsymbol{\beta}$ or team-specific parameters $\boldsymbol{\alpha}_r$. Depending on the type of parameterization, different L_1 penalty terms are set up to reduce the complexity of the model and to enable variable selection. In the case of team-specific parameters, all pairwise differences between the parameters referring to one covariate are penalized to enforce clustering of teams with equal covariate effects. Additionally, to enforce variable selection also the absolute values of all parameters can be penalized.

The method provides retrospective insights into the importance of different match-specific covariates for the success of teams in the German Bundesliga. Furthermore, it is desired to distinguish between covariates with equal effects for all teams and covariates with (clusters of) team-specific effects.

*LMU Munich

†Georg-August-University Goettingen