



## Who's the Favorite? - A Bivariate Poisson Model for the UEFA European Football Championship 2016

Andreas Groll, Gunther Schauburger  
(Institut für Statistik, LMU)

June 15, 2016, 16:00 s.t.  
Department of Statistics, Ludwigstr. 33, Room 144, Munich

A common approach in modeling soccer data is to use (univariate) Poisson regression, treating the number of goals scored by the competing teams as independent random variables given the covariate information of both teams. In this context, the information contained in the book-makers' ratings is of particular importance. It is commonly accepted that the models used by the bookmakers contain a lot of expertise as the bookmakers' profits and losses depend on the performance of their models.

In this talk we present an approach where the number of goals a team scores against a specific opponent is modeled by a joint bivariate Poisson model, including covariate information of both competing teams. The model was estimated using the R-package `gamboostLSS` (Hofner et al., 2016; Mayr et al., 2012). With `gamboostLSS` the model family of GAMLSS (Generalized Additive Models for Location, Scale and Shape) is combined with the boosting estimation technique. It allows to use multi-parametric distributions in regression models in combination with implicit variable selection.

Based on all matches from the three previous UEFA European football championships a sparse model is obtained: from a set of potential influence variables already used in Groll and Abedieh (2013) only two covariates, namely the bookmakers' odds (odds for winning the title before the tournament) and the market value were chosen. This model was then used to repeatedly simulated (100 000 times) all match outcomes of the UEFA European football championship 2016 in France, resulting in winning probabilities for all participating national teams.