

Automatic Team Selection in a Fantasy Football (Soccer) Game

Lucas da Cunha Godoy¹, Marcos Oliveira Prates², Jun Yan¹

¹ University of Connecticut, Department of Statistics

² Universidade Federal de Minas Gerais, Department of Statistics

Abbreviated abstract: Sports fantasy games, created in the first half of the twentieth century in the context of baseball, have boomed with the popularization of the internet. The idea of these games is that fans can build their own teams based on their guesses and compete against each other. The performance of the virtual teams depends on the performance of the real players. In contrast to most of the fantasy games, in Cartola FC, which is based on the 1st division of the Brazilian Championship, the users pick up their teams in every round subject to a budget restriction. This budget changes every round depending on the team performance. Aiming to select a team automatically at every round and to eliminate the subjectivity inherent to these choices, we fitted linear regressions and created an optimization algorithm to form the teams based on the predictions one round ahead constrained to the budget restriction.

Data

There is a high rotativity in the players who play in the Brazilian league mainly due to the relegation/promotion that occurs every season and to the number of players sold to teams in other countries. Although we have data available for different seasons, the fact that our model uses mainly the characteristics of the players make it hard to use the information from the previous seasons due to the rotativity of the players.

- ▶ The data are collected round by round using an API provided by the company that runs the game.
- ▶ Moving averages of the last 3 games are calculated for every variable.
- ▶ The relative strength of the team is calculated by matching the positions of each team to its opponent. For example, the relative strength of the defense of the home team is calculated by subtracting its defense strength by the opponent attacking strength.

Methods

- ▶ The models were adjusted separately for coaches, goalkeepers, and all the other positions. Additionally, the set of covariates varies for these three distinct models.
- ▶ Lagged covariates are used to make it possible to have all the information that the models need available at the time of the prediction for the next round.
- ▶ After getting the prediction, every formation is tested. For each formation we optimize the predict scores subject to the budget restriction. This optimization is made in a position-wise manner.
- ▶ We are studying different alternatives for the predictive model. For example, we intend to try Bayesian models that use the last season as a prior for the regression coefficients.
- ▶ Another option in our radar is the use of statistical learning models using as objective function a robust correlation measure between the predicted and true value to tune such models.

Findings

- ▶ In Figure 1 the coefficients for the model of all the positions (except goalkeepers and coaches) is displayed. It takes at least 15 rounds to stabilize the model coefficients.
- ▶ The Figure 2 displays the performance of two distinct ways to pick up the teams for the current season.

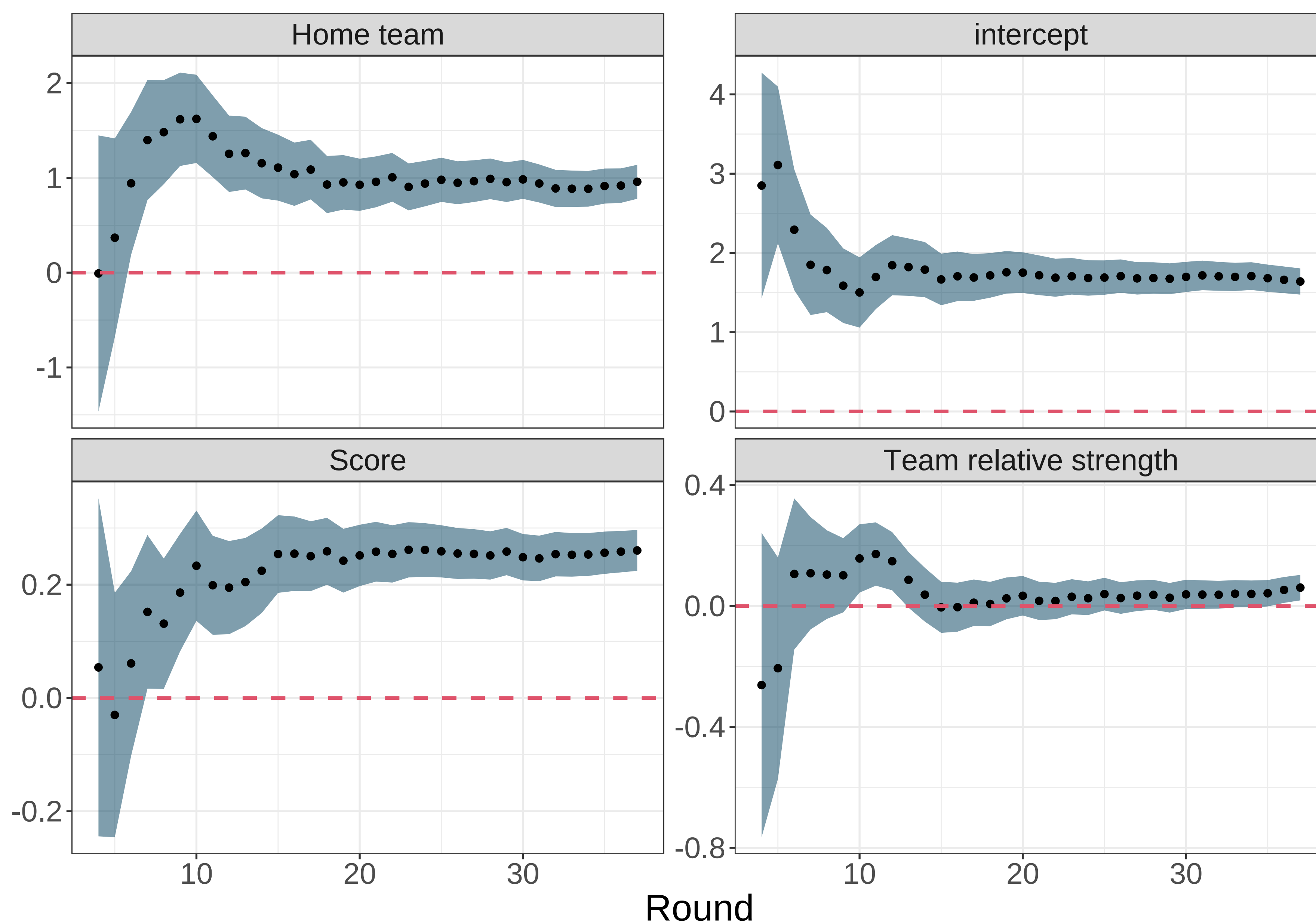


Figure 1: Estimated regression coefficients for each round of the 2019 season.

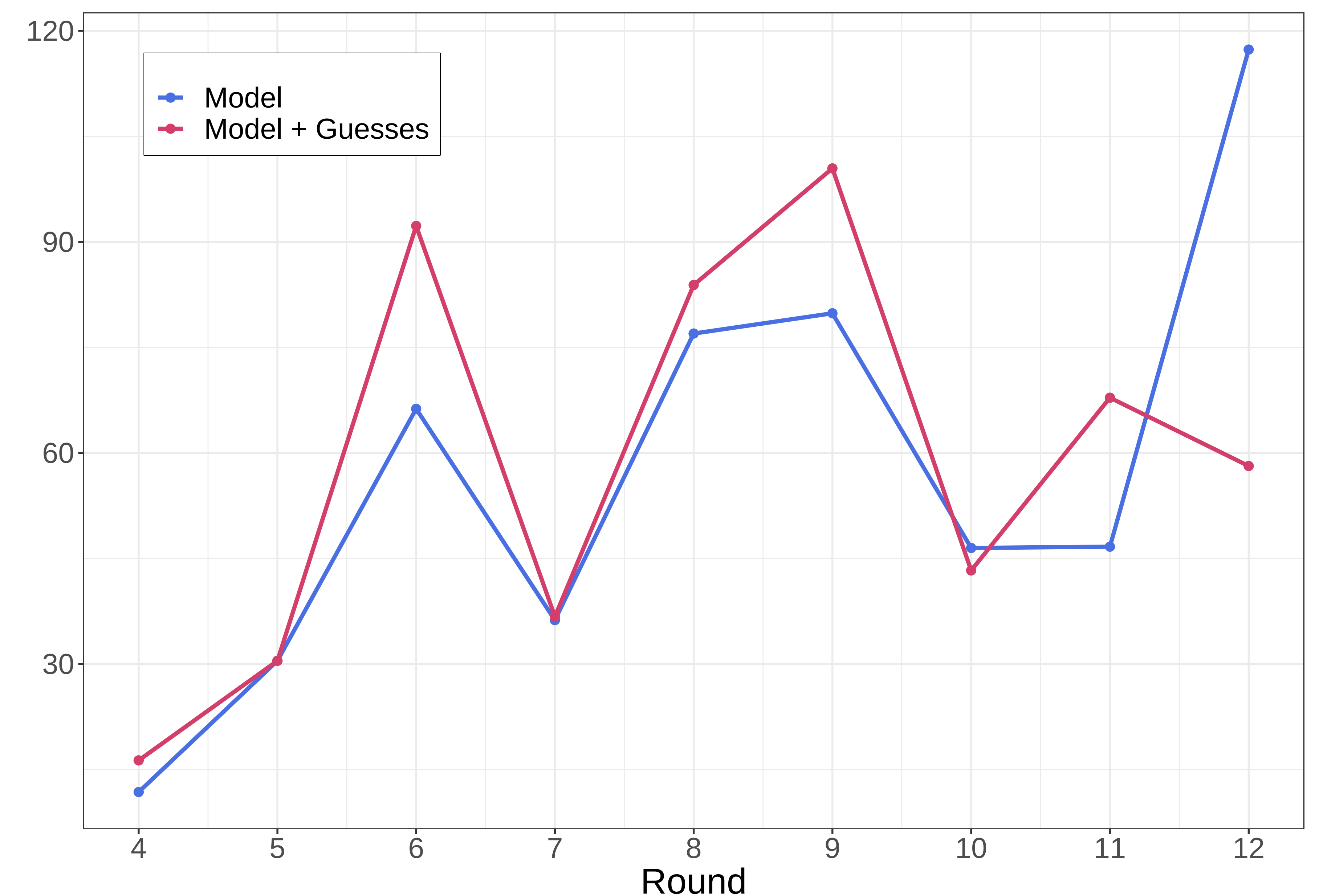


Figure 2: Performance of the selected teams for each round of the 2020 season.