Computational Statistics and Data Analysis 158 (2021) 107181

Contents lists available at ScienceDirect

Computational Statistics and Data Analysis

journal homepage: www.elsevier.com/locate/csda

Estimating robot strengths with application to selection of alliance members in FIRST robotics competitions☆

Alejandro Lim [a,*], Chin-Tsang Chiang [b,*], Jen-Chieh Teng [c,*]

[a] University of California, Berkeley, USA
[b] Institute of Applied Mathematical Sciences, National Taiwan University, Taipei, Taiwan, ROC
[c] Data Science Degree Program, National Taiwan University, Taipei, Taiwan, ROC

ARTICLE INFO

Article history:
Received 4 September 2020
Received in revised form 29 December 2020
Accepted 12 January 2021
Available online 12 February 2021

Keywords:
Accuracy
Binary regression model
Latent clusters
Offensive power rating
Winning margin power rating

ABSTRACT

Since the inception of the FIRST Robotics Competition (FRC) and its special playoff system, robotics teams have longed to appropriately quantify the strengths of their designed robots. The FRC includes a playground draft-like phase (alliance selection), arguably the most game-changing part of the competition, in which the top-8 robotics teams in a tournament based on the FRC's ranking system assess potential alliance members for the opportunity of partnering in a playoff stage. In such a three-versus-three competition, several measures and models have been used to characterize actual or relative robot strengths. However, existing models are found to have poor predictive performance due to their imprecise estimates of robot strengths caused by a small ratio of the number of observations to the number of robots. A more general regression model with latent clusters of robot strengths is, thus, proposed to enhance their predictive capacities. Two effective estimation procedures are further developed to simultaneously estimate the number of clusters, clusters of robots, and robot strengths. Meanwhile, some measures are used to assess the predictive ability of competing models, the agreement between published FRC measures of strength and model-based robot strengths of all, playoff, and FRC top-8 robots, and the agreement between FRC top-8 robots and model-based top robots. Moreover, the stability of estimated robot strengths and accuracies is investigated to determine whether the scheduled matches are excessive or insufficient. In the analysis of qualification data from the 2018 FRC Houston and Detroit championships, the predictive ability of our model is also shown to be significantly better than those of existing models. Teams who adopt the new model can now appropriately rank their preferences for playoff alliance partners with greater predictive capability than before.

1. Introduction

For Inspiration and Recognition of Science and Technology (FIRST) is an international youth organization which sponsors the FIRST Robotics Competition (FRC), a league involving three-versus-three matches, as dictated by a unique game released annually. The FIRST entry on Wikipedia (2018) provides a more detailed description of the organization and its history. Winning an individual match in the FRC requires one set of three robots, an alliance, scoring more points than

☆ The data and R codes used in this article are available at http://www.runmycode.org/companion/view/4012.

* Corresponding authors.
E-mail addresses: mogul168@berkeley.edu (A. Lim), chiangct@ntu.edu.tw (C.T. Chiang), D09948011@ntu.edu.tw (J.C. Teng).

https://doi.org/10.1016/j.csda.2021.107181
0167-9473/

an opposing alliance. Since its founding by Dean Kamen and Woodie Flowers in 1989, an increasing number of teams have designed robots according to each year's specific FRC game mechanics. The FRC has since ballooned to a pool of close to 4000 teams in 2018 playing in local and regional tournaments for a chance to qualify for the world championships of more than 800 teams. In turn, starting in 2017, FIRST divided its world championships into two championship tournaments in Houston and Detroit with about 400 robots playing in each.

Important to this article's terminology and used frequently among scouters and tournament staff alike are the competing "red" and "blue" alliances to refer to the aforementioned sets of three robots. Those interpreting raw footage of matches rely heavily on the correspondingly colored bumpers of the robots in their data collection. Additionally, a robotics "team" refers to a group of people who have constructed, managed, and maintained a "robot", which refers to the actual machine playing in a match.

With the never-changing three-versus-three format arose questions posed by many FRC strategists over the years in analyzing past tournaments and their alliance selection phases: Which robots carried their alliance, and which robots were carried by their alliance? How might one use the answers to the first questions to predict hypothetical or upcoming matches?

Over the years, FRC games have gone through a variety of changes and tasks that form a key part of competition. For 2017, tasks that earned points during a match included having the robot move across a line during the autonomous period, delivering game elements from one location to another using the robot, and having the robot climb a rope (Fig. 1). Generally, tasks will be similar to these regardless of the year, though the game elements, delivery locations, and terrain will often change each year. In addition to similar tasks, there are some fundamental aspects that have not changed from year to year. Games across the years have followed the same match timing format of 15 s of autonomous control, in which robots are pre-programmed to operate in ways that score points, followed by a period of remote control, also called the teleoperated period, lasting at least two minutes. FIRST (2017a) has provided a more detailed video guide of how points were scored for the 2017 FRC game, and also provides a similar video guide at the start of each season.

There also exist numerous regulations and loopholes to those regulations that affect the relationships of results between different tournaments. Before 2020, ideally, the same robot played in different tournaments would be expected to have the same scoring ability across those tournaments. However, analysts often consider robots of the same team from different tournaments independent from each other, since although FIRST implemented a "Stop Build Day" in which teams could no longer make significant changes to their robots, there were still loopholes that allowed for a team's robot to significantly improve in between tournaments. One such loophole allowed wealthier teams to build a secondary robot for testing new components, which could be added on during pre-tournament maintenance. As of 2020, the Stop Build Day rules have been lifted so that all teams can work on improvements more equitably in between tournaments. This development means that in future years, one can predict the performance difference of a robot between tournaments to be even more drastic than in recent history, thus reducing the reliability of previous tournament data in predicting the performance of a robot in future tournaments. Because of these developments and the history of inconsistency between tournaments, for the purposes of modeling, the same team's entry played in different tournaments will be considered a completely different robot. The model presented in this article is, thus, limited to the data from the qualification rounds to predict the outcomes of the playoff rounds from the same tournament. It must also be noted that most maintenance that occurs in between rounds does not have a significant additional positive impact on the robot's performance, since most of the scoring ability is inherently designed outside of competitive play.

The FRC's rules provide a structure with many features analogous to traditional sports. A kickoff event in January announces the rules of that year's game and signifies the beginning of the "build season". Teams often organize scrimmages or collaboratively test prototypes during this time in a fashion similar to roster formation during the MLB's Spring Training, other preseasons, or even free agent workouts. Build season is followed by the competition season where good tournament results count toward advancement to increasingly higher levels of gameplay, most notably toward the aforementioned championship tournaments.

As of the 2018 competitions, each championship site is composed of six divisions. Each division runs a mini-tournament to determine a division champion alliance. The six division champion alliances further advance to the Einstein Division, a round-robin group whose top two alliances play a best-of-three to decide a site champion alliance. In some years, each site champion alliance plays the other at another gathering called the "Festival of Champions". This scheme somewhat parallels the Major League Baseball's (MLB) National and American Leagues before the introduction of interleague play in that championship sites, championship divisions, and regions for qualifying for the world championships segregated teams from all over the world into their own pools of interaction, with the Festival of Champions serving as a counterpart to the MLB's World Series.

As in all tournaments throughout a season, matches in each division mini-tournament are divided into a qualification stage and a playoff stage, which are similar to a microcosm of a regular season and a postseason in major sports leagues on the scale of only two days and with far more teams. In the qualification stage, teams' robots are assigned by a predetermined algorithm into matches of six split into three in a blue alliance against another three in a red alliance. Deshpande and Jensen (2016) found similar types of data in the National Basketball Association (NBA), where the contributions of players are assessed through home team scores and away team scores at different shifts, which are defined to be periods of play between substitutions when ten players on the court is unchanged, during a match. Such a concept is also applicable in the National Hockey League (NHL).

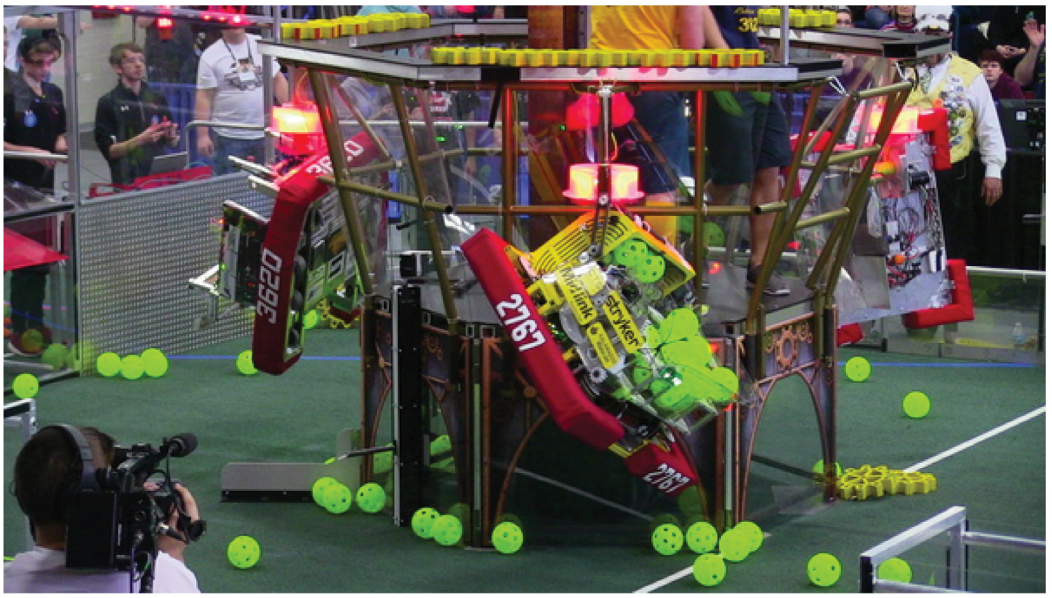

Fig. 1. Three robots on the red alliance demonstrate their ability to climb a rope and hold wiffle balls. The yellow gears were delivered from an angled slot or from the ground by the robots to be used in turning a crank on the top ledge. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)
Source: https://www.thebluealliance.com/team/2767/2017.

Within the context of an individual game, the FRC also shares many similarities with traditional sports. Basketball is a game that could be broken into separate units such as possessions, in which one team has control of the ball and attempts to score by putting the ball in a basket, while the other team defends said basket. The above description defines a possession with a scoring method, a defense, and an offense. The FRC's analog to the NBA's possession is a cycling feature in which a robot may use the same method or set of actions to score repeatedly over the course of a match. Like the structure of a basketball possession, each of these cycles involves an offensive unit attempting to score with a specific action, and, sometimes, a defensive unit, which seeks to deter or defend from opponent scoring attempts. Robots on the same alliance may work in tandem within each cycle or in parallel, often running multiple independent cycles with the hopes of maximizing scoring. In both cases, a successful score for the offense is seen positively for offensive players and negatively for defensive players. That the FRC allows the possibility of simultaneous, independent cycles should not lessen the similarity between a cycle and a possession, since each cycle may be modeled the same way. The relation between cycles in the FRC and possessions in the NBA allows individual contributions of robots in a FIRST match to be analyzed like the individual contributions of players. However, unlike the FRC, players in the NBA are not randomly assigned to shifts, and if a shift in the NBA is comparable to an individual match in the FRC, for any given NBA shift within a game, players of different teams cannot be teammates in traditional sports, whereas in the FRC, former teammates often end up facing each other in later matches on the same day.

Especially worth mentioning is that for the FRC, since matches and playing field setup are standard without significant opportunity for fan interaction, there is no need to take into account a home-court advantage effect. A defensive component is also unnecessary in our model formulation. Since the designed robots are memory-less, scores of different matches are reasonably assumed to be mutually independent. There also exists the case in traditional sports where certain players may have so much synergy that a team simply does not provide the opportunity, and thus the data, for the hypothetical shifts where said players are separated. For the qualification stage of an FRC tournament, a predetermined schedule algorithm, as detailed in the "match assignment" section of the 2017 game manual (FIRST, 2017b) provides more opportunities for different synergies to be tested and for the separation of two or three well partnered robots in the available data. Effectively, the FRC tournament design eliminates the analogous power of the coaches of traditional sports to influence the model's training data in less than ideal ways. It should be noted, however, that even though minute differences exist in the updated wording of the algorithm between 2017's game and games after 2017, the algorithm itself has not changed. Although the scheduling algorithm is not truly random, it does eliminate many potential schedules that may be obviously unbalanced. Thus, the way the qualification round schedules are assigned only helps models account for irregularities in rating robots. More importantly, included in the eliminated potential schedules are those that repeat match combinations. This control diversifies the head-to-head matchups between potential alliances, but since the scheduling algorithm forces uniqueness in matchmaking within a tournament and that the same team's robot is considered a completely different entity between tournaments, every single qualification match across an entire season

is unique and unrepeatable within qualification. This makes simulating new matches very prone to error and hard to control, since any simulation of six robots in the same alliances would rely on only one pair of a red and blue score, if it exists. In fact, the only place where matches of the same six robots separated into the same alliances is possible is during the best-of -3 match-ups in the playoffs, and even that is not guaranteed to produce replicated matches, since an alliance of four robots in the playoffs can opt to swap in its substitute robot between matches.

It has been observed that within an alliance for a specific match, there is at least some robot on the winning alliance that contributed, possibly more than the others, to the win and at least some robot on the losing alliance that did not pull its weight, possibly more than the others, in the loss. For example, Team 254 of the 2018 Hopper Division went undefeated in its qualification matches, and Team 4775 of the same division failed to win a match in qualification per The Blue Alliance's online database (The Blue Alliance, 2018). These, together, provides a motivation to estimate robot strengths to determine which robots are contributing the most to wins and losses. We aim in this article to develop a model that utilizes the estimated strengths of individual robots to predict the outcome of any match.

Official ranking of FRC teams within the qualification stage involves a system of ranking points (RP). The kickoff event's revealing of game rules introduces two in-game objectives, the completion of which is rewarded with one RP each. Additionally, winning a match will net a team two RPs while losing produces no additional RP. Pre-playoff rankings are determined according to the total RP earned in the qualification stage with the top-8 teams in terms of RP being guaranteed a spot in the playoffs. These top-8 teams are given the ability to form playoff alliances with an additional three members for the playoff stage, one of which serves as a substitute, preserving the three-versus-three game structure. When there are ties within the RP system, the average score (AS) measure is used to break RP ties.

Unlike the qualification stage, playoff stage alliances are determined by teams during the alliance selection phase. It is during the alliance selection phase that the pool of teams within a particular tournament draft themselves into playoff alliances that they each believe have the best chances of winning in the playoff stage. It is after this phase that the new, team-selected alliances take full control of their own results, independent of randomness in having carried or being carried in the qualification stage. The alliance selection phase, which takes place immediately after the end of qualification and directly affecting the upcoming playoff, also marks the transition between when RP is king (qualification) and when in-game scores become the ultimate goal (playoff). These reasons are why alliance selection is the most game-changing and consequential portion of an entire tournament. Subsequently, there are some teams that design their robots specifically to complete the in-game objectives, important to obtaining a higher RP ranking, with the hopes of guaranteeing a spot in the playoffs and having more control over alliance selection. Teams that choose to implement this design strategy sometimes sacrifice scoring ability, which is more important during the playoff stage. However, sometimes this tradeoff for alliance selection control is understandable, since such teams usually invest more than others into scouting and analysis of the competition and would thus have an information advantage in making alliance selections.

The alliance selection period is like the free agency period of the NBA. This is when players and managements, aside from in-season trades, negotiate the bulk of partnerships in, for non-rebuilding teams, hopes of capturing the best odds for the following season. It is important to note that unlike traditional sports leagues, the FRC does not have an analog to team-rebuilds within a tournament, so all alliance selections must be done with only the end goal of a championship in mind. Similarly, robotics teams, will negotiate with each other during the alliance selection phase to form the best alliance of robots for the playoff rounds. Synergy comes into play within both the free agency period and alliance selection, as management of an NBA team would not be looking to hypothetically field a team of five centers, nor would a FIRST team necessarily look to form a partnership with a team whose robot is optimized for the same scoring method. The best alliances tend to have each robot perform a different job, so as to maximize scoring opportunity, or minimize the other alliance's scoring opportunity if one of the jobs is general defense, depending on that year's game. Thus, NBA management and FRC strategists are comparable in that they both seek well-roundedness, as well as saturated ability in their quests for championships. Other traditional sports' free agencies also exhibit the same desires and goals from teams.

In the playoff stage, the eight alliances play three of their four robots in each match in a best of three matches knockout manner until a champion is declared. As the FRC continues to progress, so have questions from participants on improving the quality of alliance selection and decision making. Since the goal of playoff matches is actual points rather than RP, many decide not to rely on the RP system to make decisions regarding the alliance selection. Teams throughout the years have created new measures in order to assess actual or relative robot strengths. Some widely used measures in robotics forums (e.g. Weingart, 2006; Law, 2008; Gardner, 2015; Fang, 2017) include the offensive power rating (OPR), the winning margin power rating (WMPR), and the calculated contribution to winning margin (CCWM), among others. In one such investigation of these measures, using simulated and past FRC and FTC tournaments, Gardner (2015) further analyzed the performance of these measures. However, the OPR and WMPR models were found to have poor predictive performance due to their imprecise estimates caused by a small ratio of the number of observations to the number of robots in the considered data. Furthermore, the CCWM model was shown to be a special case of the WMPR model and indicated to be meaningless in our application.

In the 2018 FRC Houston and Detroit championships, there were 112 to 114 matches involving 67 to 68 robots for divisions in the qualification stage. The corresponding ratios of the number of observations to the number of robots range from 3.28 to 3.40 for the OPR model and from 1.64 to 1.70 for the WMPR model. The analysis of such paired comparison data also reveals no strong evidence to support the use of the WMPR model. These considerations motivate us to explore some possible avenues to enhance their predictive capacities. With the consideration of latent clusters of robot strengths,

the proposed model can be regarded as the dimension reduction of the parameter spaces in the OPR and WMPR models. One crucial issue is how to estimate the number of clusters, clusters of robots, and robot strengths. The major aim of our proposal is to assess robot strengths in a more accurate and precise manner and help highly ranked teams assess potential alliance members for the opportunity of partnering in the playoff stage. To the best of our knowledge, there is still no research devoted to studying the latent clusters of individual strengths in team games.

The rest of this article is organized as follows. Section 2 outlines existing measures and models for robot strengths. In Section 3, we propose a more general regression model with latent clusters of robot strengths, develop two effective estimation procedures, and present some agreement and stability measures. In Section 4, our proposal is applied to the 2018 FRC Houston and Detroit championships. Conclusion and discussion are also given in Section 5.

2. Existing measures and models for robot strengths

Throughout history, assessments for individual strengths have always intrigued analysts of team sports, e.g., individual stats of NBA players in basketball. This is no different in the FRC as in the last 25 years, as the FRC game designs contain features and components analogous to those in other traditional sports. Both FIRST and FRC teams have established their own systems for rating competing robots. In this section, we concisely introduce and compare the OPR, CCWM, and WMPR measures and models for robot strengths. Different from the OPR, CCWM, and WMPR models for the match outcome in the literature, a more general semi-parametric formulation is further presented in this study. Moreover, we review similar measures and models for individual strengths in other team games.

2.1. Data and notations

Let K and M stand for the number of robots and the number of matches, respectively, in a division. In each match s, three robots forming the blue alliance, B_s, go against three other robots forming the red alliance, R_s. In the qualification stage, the first $\lceil \frac{1\times K}{6} \rceil$ matches are designed to ensure that each robot plays at least one match, the succeeding $\lceil \frac{1\times K}{6} \rceil$ matches are designed to ensure that each robot plays at least two matches, and in total, $M = \lceil \frac{m_0 \times K}{6} \rceil$ matches are designed to ensure that each robot plays at least m_0 matches, where $\lceil\rceil$ is the ceiling function. For example, in the Carver division of the 2018 Houston championship, $K = 68$, $M = 114$, and $m_0 = 10$. With a total of 114 matches, each robot played at least ten matches. Since exactly four robots played eleven matches, per the rules as specified in the game manual (FIRST, 2017b), the third match of each robot that played eleven games was not considered in their rankings, so that all robots have exactly ten matches worth of opportunity to perform.

To simplify the presentation, let Y_s^{B} and Y_s^{R} stand for the scores of B_s and R_s, respectively; $\beta_1, \ldots, \beta_K$ the strengths (or the relative strengths with the constraint $\sum_{i=1}^{K} \beta_i = 0$) of robots with the corresponding IDs $1, \ldots, K$; and *i.i.d.* the abbreviation of independent and identically distributed. The super-index in β_i's and their estimators is further used to distinguish different measures. For the linear model formulation of existing models, we also define the following notations:

$$D_s = I\big(Y_s^{\mathrm{R}} - Y_s^{\mathrm{B}} > 0\big), X_{si}^{\mathrm{B}} = I(i \in \mathrm{B}_s), X_{si}^{\mathrm{R}} = I(i \in \mathrm{R}_s),$$
$$Y^{(t)} = \big(Y_1^{(t)}, \ldots, Y_M^{(t)}\big)^\top, \text{ and } X^{(t)} = \big(X_{si}^{(t)}\big) = \big(X_1^{(t)}, \ldots, X_M^{(t)}\big)^\top = \big(X_{(1)}^{(t)}, \ldots, X_{(K)}^{(t)}\big),$$

where $I(\cdot)$ is the indicator function, $X_s^{(t)} = (X_{s1}^{(t)}, \ldots, X_{sK}^{(t)})^\top$ and $X_{(i)}^{(t)} = (X_{1i}^{(t)}, \ldots, X_{Mi}^{(t)})^\top$ represent the covariate information of match s and robot i, respectively, $s = 1, \ldots, M$, $i = 1, \ldots, K$, and (t) can be either B or R.

2.2. Conceptual models

In application, the AS, the average score, is a simple way to assess robot strengths. The AS of a robot is calculated by adding up the alliance scores in the matches played and dividing by three times of the number of matches played. Under this system, the strength of robot i is estimated by

$$\hat{\beta}_i = \frac{\sum_{s=1}^{M}\big(X_{si}^{\mathrm{B}} Y_s^{\mathrm{B}} + X_{si}^{\mathrm{R}} Y_s^{\mathrm{R}}\big)}{3\sum_{s=1}^{M}\big(X_{si}^{\mathrm{B}} + X_{si}^{\mathrm{R}}\big)}, i = 1, \ldots, K. \tag{1}$$

Different from the plus–minus statistic, which was first adopted by the NHL's Montreal Canadiens in the 1950s, the AS is analogous to goals for or points for, instead of score differential, divided by the number of players. Same as the drawback of the plus–minus statistic, obviously strong (weak) robots may be underrated (overrated) due to being paired with relatively weak (strong) robots, albeit robots are randomly assigned to matches. Even with a large enough number of matches, the AS is still not a good representation for robot strengths.

As detailed by Fang (2017), in 2004, Karthik Kanagasabapthy of the FRC Team 1114 created a measure, which was termed the calculated contribution, to assess the contribution of a robot to an alliance score. He initiated work on the OPR measure, which characterizes the alliance score on the sum of contribution strengths of the alliance's robots, and estimated robot strengths by the least squares solution of systems of equations. The details of the computation were

further explained in a post by Weingart (2006), who first called the calculated contribution as the OPR. In the following OPR model, the blue alliance score and the red alliance score in each match are formulated in the same way:

$$Y_s^{\mathrm{R}} = \sum_{\{i\in \mathrm{R}_s\}} \beta_i + \varepsilon_s^{\mathrm{R}} \text{ and } Y_s^{\mathrm{B}} = \sum_{\{i\in \mathrm{B}_s\}} \beta_i + \varepsilon_s^{\mathrm{B}}, s = 1, \ldots, M, \tag{2}$$

where $\varepsilon_1^{\mathrm{R}}, \ldots, \varepsilon_M^{\mathrm{R}}, \varepsilon_1^{\mathrm{B}}, \ldots, \varepsilon_M^{\mathrm{B}}$ are *i.i.d.* with mean zero and variance σ^2. As we can see, the above formulation is similar to that from Macdonald (2011) for NHL players except that the OPR model does not consider a home-court advantage effect and a defensive component. Under the assumption of independent errors, $2M$ observations are used in the estimation of $\beta_1, \ldots, \beta_K$. While such a model formulation allows for the decomposition of expected alliance scores into individual robot strengths in a logical manner, it completely ignores actions of robots in the opposing alliance. As we can see, the independence between $\varepsilon_s^{\mathrm{R}}$ and $\varepsilon_s^{\mathrm{B}}$, $s = 1, \ldots, M$, is unrealistic in practice. Furthermore, robots in blue and red alliances might be influenced by a common unobserved factor (e.g. environmental obstacles), say Z_s, in match s such that $E[Z_s|\{i : i \in \mathrm{B_s} \text{ or } \mathrm{R_s}\}] = E[Z_s] = \beta_0 \neq 0$, $s = 1, \ldots, M$. To achieve this more realistic interpretation, the OPR model can be modified as

$$Y_s^{\mathrm{R}} = \beta_0 + \sum_{\{i\in \mathrm{R}_s\}} \beta_i + \varepsilon_s^{\mathrm{R}} \text{ and } Y_s^{\mathrm{B}} = \beta_0 + \sum_{\{i\in \mathrm{B}_s\}} \beta_i + \varepsilon_s^{\mathrm{B}}, s = 1, \ldots, M, \tag{3}$$

where the intercept β_0 can be interpreted as the average of all robot strengths. Furthermore, the error terms $\varepsilon_s^{\mathrm{R}}$ and $\varepsilon_s^{\mathrm{B}}$ cannot be assumed to be mutually independent, $s = 1, \ldots, M$. It is notable that the resulting estimators of the regression coefficients in model (2) are biased estimators of those in model (3).

Another commonly used assessment for robot strengths is the WMPR measure from Gardner (2015). This measure accounts for the effects of robots on the opposing alliance. With this consideration, the difference in scores between alliances in the WMPR model is formulated as

$$Y_s^{\mathrm{R}} - Y_s^{\mathrm{B}} = \sum_{\{i\in \mathrm{R}_s\}} \beta_i - \sum_{\{i\in \mathrm{B}_s\}} \beta_i + \varepsilon_s, s = 1, \ldots, M, \tag{4}$$

where $\varepsilon_1, \ldots, \varepsilon_M$ are *i.i.d.* with mean zero and variance σ^2. In basketball, soccer, volleyball, and esport games, some research works such as Rosenbaum (2004), Macdonald (2011), Schuckers et al. (2011), Sæbø and Hvattum (2015), Hass and Craig (2018), Hvattum (2019), and Clark et al. (2020), among others, used the adjusted plus/minus rating (APMR), which is similar to the WMPR, to assess the contribution to net scores per possession by each player relative to all other players. In contrast with the WMPR model, the APMR model takes into account a home-court advantage and clutch time/garbage time play, typical of many team sports. In traditional sports, this model formulation has been shown to be useful for paired comparison data. A more general condition is imposed on $\varepsilon_1, \ldots, \varepsilon_M$, albeit the OPR model in (2) or (3) also leads to

$$Y_s^{\mathrm{R}} - Y_s^{\mathrm{B}} = \sum_{\{i\in \mathrm{R}_s\}} \beta_i - \sum_{\{i\in \mathrm{B}_s\}} \beta_i + \left(\varepsilon_s^{\mathrm{R}} - \varepsilon_s^{\mathrm{B}}\right), s = 1, \ldots, M, \tag{5}$$

where $(\varepsilon_1^{\mathrm{R}} - \varepsilon_1^{\mathrm{B}}), \ldots, (\varepsilon_M^{\mathrm{R}} - \varepsilon_M^{\mathrm{B}})$ are *i.i.d.* with mean zero and variance $E[(\varepsilon_1^{\mathrm{R}} - \varepsilon_1^{\mathrm{B}})^2]$. However, only M observations are used in the estimation of robot strengths.

By taking into account the influence of the opposing alliance score, Law (2008) showed a CCWM model of the following form:

$$Y_s^{\mathrm{R}} - Y_s^{\mathrm{B}} = \sum_{\{i\in \mathrm{R}_s\}} \beta_i + \varepsilon_s^{\mathrm{R}} \text{ and } Y_s^{\mathrm{B}} - Y_s^{\mathrm{R}} = \sum_{\{i\in \mathrm{B}_s\}} \beta_i + \varepsilon_s^{\mathrm{B}}, s = 1, \ldots, M, \tag{6}$$

where $\varepsilon_1^{\mathrm{R}}, \ldots, \varepsilon_M^{\mathrm{R}}, \varepsilon_1^{\mathrm{B}}, \ldots, \varepsilon_M^{\mathrm{B}}$ are *i.i.d.* with mean zero and variance σ^2. Clearly, the CCWM model formulates the net effect of the opposing alliance score on the reference alliance score. The contribution of a robot in its participated match is further explained by the winning margin. However, by the equality $E[Y_s^{\mathrm{R}} - Y_s^{\mathrm{B}}] = -E[Y_s^{\mathrm{B}} - Y_s^{\mathrm{R}}]$, we have $\sum_{\{i\in \mathrm{R}_s\}} \beta_i = -\sum_{\{i\in \mathrm{B}_s\}} \beta_i$, $s = 1, \ldots, M$, which leads to $\beta_1 = \cdots = \beta_K = 0$ for $M \geq K$ and $\varepsilon_s^{\mathrm{R}} = -\varepsilon_s^{\mathrm{B}}, s = 1, \ldots, M$. As a result,

$$Y_s^{\mathrm{R}} - Y_s^{\mathrm{B}} = \varepsilon_s, s = 1, \ldots, M, \tag{7}$$

where $\varepsilon_1, \ldots, \varepsilon_M$ are *i.i.d.* with mean zero and variance σ^2. This fact indicates that the CCWM model is meaningless in our application. Obviously, the model formulation of the CCWM model in (7) is a special case of that in (4) with $\beta_1 = \cdots = \beta_K = 0$. For this reason, the CCWM model will not be studied in the rest of this article.

Remark 1. For the defensive strengths of robots, there were some proposals to take into account the defensive power rating (DPR). It was shown by Gardner (2015) that estimated robot strengths of the DPR model can be expressed as the difference of those of the OPR and CCWM models. The author further introduced other measures such as combined power rating, mixture-based ether power rating, and some related simultaneous measures. However, in the setup of the FRC system, these measures are inappropriate to characterize robot strengths. For this reason, we do not explore their properties and extensions in this article. □

2.3. Comparisons to measures in other team games

Wins above replacement (WAR), as described by Slowinski (2010), and value over replacement player (VORP), detailed by Woolner (2001a,b), are statistics used in baseball to express individual contributions of a player to his team. WAR measures this by wins in a 162-game season, while VORP measures offensive contributions of a player through runs created and pitching contributions through allowed run averages. These statistics are comparable to the different individual contribution statistics analyzed in this article. Out of these two statistics, VORP is more similar to the currently available robotics statistics, since the lower number of matches in qualification, which does not usually pass 11 per robot for the 2018 championships, means there is not enough data to fit an accurate WAR-like measure.

While the APMR model seems to gain an advantage in describing the basketball, ice hockey, soccer, volleyball, and esport games, the context is quite different from robotics competitions. Furthermore, Ilardi and Barzilai (2008) incorporated the role of each player as offensive or defensive into the APMR model. Based on the difference between the expected and observed outcomes, Schuckers et al. (2011) proposed an adjusted plus–minus probability model. In a robotics tournament, the ratio of the number of observations to the number of robots in each division is under two for the WMPR model, while the ratio of the number of possessions to the number of NBA players in a season is about sixteen (see TeamRankings (2019) for the 2018–2019 season) for the APMR model. This limitation in the WMPR model usually produces poor estimation of robot strengths and poor prediction of future outcomes. An important task of our research is, thus, to develop a better predictive model.

2.4. Linear model formulation

It follows from (2) and (4) that the OPR and WMPR models can be expressed as

$$Y = X\beta + \varepsilon, \tag{8}$$

where $\beta = \left(\beta_1, \ldots, \beta_K\right)^T$ and ε has mean vector $0_{\bar{M}\times 1}$ and covariance matrix $\sigma^2 I_{\bar{M}}$. In terms of such a linear model formulation, we have

1. $Y = \begin{pmatrix} Y^{\mathrm{R}} \\ Y^{\mathrm{B}} \end{pmatrix}$, $X = \begin{pmatrix} X^{\mathrm{R}} \\ X^{\mathrm{B}} \end{pmatrix}$, and $\varepsilon = \left(\varepsilon_1^{\mathrm{R}}, \ldots, \varepsilon_M^{\mathrm{R}}, \varepsilon_1^{\mathrm{B}}, \ldots, \varepsilon_M^{\mathrm{B}}\right)^\top$ with $\bar{M} = 2M$ in the OPR model; and
2. $Y = Y^{\mathrm{R}} - Y^{\mathrm{B}}$, $X = X^{\mathrm{R}} - X^{\mathrm{B}}$, and $\varepsilon = (\varepsilon_1, \ldots, \varepsilon_M)^T$ with $\bar{M} = M$ in the WMPR model.

Owing to the linear dependence of $\left(X_{(i)}^{\mathrm{R}} - X_{(i)}^{\mathrm{B}}\right)$'s (i.e. $\sum_{i=1}^{K}\left(X_{(i)}^{\mathrm{R}} - X_{(i)}^{\mathrm{B}}\right) = 0$) in the WMPR model, the constraint $\sum_{i=1}^{K}\beta_i = 0$, which was also adopted by Cattelan et al. (2013) for the Bradley–Terry specification in Bradley and Terry (1952) and Bradley (1953), is further imposed to solve the identifiability of β. As a result, the coefficient β_i, compared to β_K, is explained as the relative strength of robot i, $i = 1, \ldots, K-1$. With this constraint, model (8) can be rewritten as

$$Y = \bar{X}\beta^* + \varepsilon, \tag{9}$$

where $\bar{X}$ is a $M \times (K-1)$ matrix with the ith column being $X_{(i)} - X_{(K)}$ and $\beta^* = \left(\beta_1, \ldots, \beta_{K-1}\right)^\top$ with $\beta_K = -\sum_{i=1}^{K-1}\beta_i$. In the context of regression analysis, β (or β^*) is naturally estimated by the least squares estimator (LSE), say $\widehat{\beta}$ (or $\widehat{\beta}^*$), under the Gauss–Markov conditions.

Remark 2. Under the WMPR model, Gardner (2015) proposed an estimator of β by solving the pseudo-inverse solution of the corresponding estimating equations of the sum of squares. Albeit lacking an explanation for the resulting estimator, the estimated or predicted score of a match is unique. In application, the constraint $\sum_{i=1}^{K}\beta_i^2 = 1$ can also be imposed to obtain estimators of relative robot strengths. However, there is no simple formulation for the corresponding LSE of β. □

Remark 3. Like the development of a plus/minus rating system by Feamhead and Taylor (2011), the robot strengths β_i's in the WMPR model can also be formulated as *i.i.d.* normal random variables with mean zero and variance σ^2. Together with the normality assumption on ε and the independence between β and ε, the resulting predictor of β has been shown by Fahrmeir and Tutz (2001) to be a ridge estimator, which is the same as that in Sill (2010), with the regularization parameter σ^2/σ_0^2 in such a Bayesian framework. Since the ridge regression is mainly used to combat multicollinearity of covariates, its explanation is different from that of the introduced Bayes estimator. Furthermore, in most existing paired comparisons models, this problem was solved by imposing the constraint $\sum_{i=1}^{K}\beta_i = 0$. In our application to the 2018 FRC championships, it is also shown that the WMPR model and its Bayesian formulation, which is hereinafter referred to as the WMPRR model, have comparable performance in prediction. □

2.5. Binary regression models

It is implied by the OPR and WMPR models that the corresponding conditional probabilities of a match outcome have the following form:

$$P\left(D_s = 1 | X_s^{\mathrm{B}} = x_s^{\mathrm{B}}, X_s^{\mathrm{R}} = x_s^{\mathrm{R}}\right) = 1 - F\left(-(x_s^{\mathrm{R}} - x_s^{\mathrm{B}})^\top\beta\right), s = 1, \ldots, M, \tag{10}$$

where $F(\cdot)$ is an unknown distribution function. For the OPR model, $F(\cdot)$ is a distribution function of $(\varepsilon_s^{\mathrm{R}} - \varepsilon_s^{\mathrm{B}})$'s. Although the conditional probability in (10) can also be derived from model (3), the LSEs of the regression coefficients in both models and the resulting nonparametric estimators of $F(\cdot)$ will be completely different. As for the WMPR model, $F(\cdot)$ is a distribution function of ε_i's.

Let $(Y_0^{\mathrm{B}}, Y_0^{\mathrm{R}}, X_0^{\mathrm{B}}, X_0^{\mathrm{R}})$ stand for a future run with $D_0 = I\big(Y_0^{\mathrm{R}} - Y_0^{\mathrm{B}} > 0\big)$ and $(x_0^{\mathrm{B}}, x_0^{\mathrm{R}})$ being a realization of $(X_0^{\mathrm{B}}, X_0^{\mathrm{R}})$. In practical implementation, we will conclude that

$$\{Y_0^{\mathrm{R}} > Y_0^{\mathrm{B}}\} \text{ if } F\big(-(x_0^{\mathrm{R}} - x_0^{\mathrm{B}})^{\top}\beta\big) < 0.5 \text{ and } \{Y_0^{\mathrm{R}} \le Y_0^{\mathrm{B}}\} \text{ otherwise.} \tag{11}$$

Especially worth mentioning is that we do not assume any structure on $F(\cdot)$. For a strict monotonicity of $F(u)$ with $F(0) = 0.5$, $F\big(-(x_0^{\mathrm{R}} - x_0^{\mathrm{B}})^{\top}\beta\big) < 0.5$ is equivalent to $\big(x_0^{\mathrm{R}} - x_0^{\mathrm{B}}\big)^{\top}\beta > 0$. Special cases of $F(u)$ for paired comparison data include the logistic distribution function and the standard normal distribution function, which lead to the Bradley–Terry model in Bradley and Terry (1952) and Bradley (1953) and the Thurstone–Mosteller model in Thurstone (1927) and Mosteller (1951), respectively. In various sports and games, the Bradley–Terry model has been widely used to rate players. It includes the studies of Cattelan et al. (2013) in basketball and football, Elo (1978) in chess and other sports, Islam et al. (2017) in cricket, Chen et al. (2017) and Weng and Lin (2011) in online multi-player games, and Huang et al. (2008) in bridge. With the specifications of the complementary logistic function for the "update function shape" and the logistic model for the margin of victory in the Elo-type rating systems, Aldous (2017) and Kovalchik (2020), respectively, showed that the resulting models are related to the Bradley–Terry model.

3. Latent clusters of robot strengths

A more general model formulation, compared to existing models, is used to characterize latent clusters of robot strengths. Two effective estimation procedures are further developed for the proposed regression model. Moreover, some measures for the predictive ability of models, the agreement related to FRC ratings and model-based robot strengths, and the stability of estimated robot strengths and accuracies are presented in this section.

3.1. Model extensions

In Figs. 2 and 3, we observe a clustering feature in the estimated robot strengths from the OPR and WMPR models for these two championship tournaments. With this finding, both models are extended to a more general formulation with a clustering feature of robot strengths, i.e.

$$\beta = \left(\beta_{g_1}^{c_0}, \ldots, \beta_{g_K}^{c_0}\right)^{\top},$$

where c_0 is an unknown number of clusters and $g_i \in \{1, \ldots, c_0\}$ is the corresponding cluster of robot i, $i = 1, \ldots, K$, with $g = (g_1, \ldots, g_K)^{\top}$. In addition to reducing the number of parameters in the OPR and WMPR models, we do not assume any particular distribution for the underlying distributions of the difference in scores and the match outcome (win/loss). In fact, the viewpoint of grouping players into some skill levels has been adopted by the United States Chess Federation (USCF) and the National Wheelchair Basketball Association (NWBA). By using a rating system as defined by Elo (1978), which is used to assess the relative skill levels of players, the USCF classifies chess players into thirteen groups: class A, ..., class J, expert or candidate master, national master, and senior master. Based on the player's physical capacity to execute fundamental basketball movements, the NWBA groups players into one of eight categories (1, 1.5, ..., 4, and 4.5). However, these classification criteria are slightly subjective in practical implementation. By incorporating the information of $\beta_i = \beta_{g_i}^{c_0}$ with $g_i \in \{1, \ldots, c_0\}$, $i = 1, \ldots, K$, into the model formulation $Y = \beta_1 X_{(1)} + \cdots + \beta_K X_{(K)} + \varepsilon$ in (8), the covariates $X_{(i)}$'s sharing the same coefficient, say $\beta_j^{c_0}$, can be further summed together to produce a new covariate $X_{(j)}^{c_0}$, $j = 1, \ldots, c_0$. It follows that Y can be expressed as $\beta_1^{c_0} X_{(1)}^{c_0} + \cdots + \beta_{c_0}^{c_0} X_{(c_0)}^{c_0} + \varepsilon$ and model (8) is an overparameterized version of model (12) for $K > c_0$.

Based on theoretical and practical considerations, an extended linear regression model is, thus, proposed as follows:

$$Y = X^{c_0}\beta^{c_0} + \varepsilon, \tag{12}$$

where $\beta^{c_0} = (\beta_1^{c_0}, \ldots, \beta_{c_0}^{c_0})^{\top}$, ε has mean vector $0_{\bar{M}\times 1}$ and covariance matrix $\sigma_0^2 I_{\bar{M}}$, and $X^{c_0} = (X_1^{c_0}, \ldots, X_M^{c_0})^{\top} = (X_{(1)}^{c_0}, \ldots, X_{(c_0)}^{c_0})$ is a designed covariate matrix of X^{Bc_0} and X^{Rc_0} with $X_{(j)}^{Bc_0}$ and $X_{(j)}^{Rc_0}$ being defined as $X_{(j)}^{c_0}$ based on $X_{(i)}^{B}$'s and $X_{(i)}^{R}$'s, respectively. Note that model (8) is our special case with $c_0 = K$ and is over-parameterized when c_0 is smaller than K. In addition to generalize the model formulation of robot strengths in (8), we do not assume any particular distribution for the error term ε. Since the ratio of the number of matches to the number of robots in each division is rather small, our study provides a possible avenue to enhance the predictive capacities of existing models. Same with the derivation in (10), we have

$$P\big(D_s = 1 | X_s^{\mathrm{B}} = x_s^{\mathrm{B}}, X_s^{\mathrm{R}} = x_s^{\mathrm{R}}\big) = 1 - F\big(-(x_s^{Rc_0} - x_s^{Bc_0})^T\beta^{c_0}\big), s = 1, \ldots, M, \tag{13}$$

for the OPR model with latent clusters of robot strengths (OPRC) and the WMPR model with latent clusters of robot strengths (WMPRC).

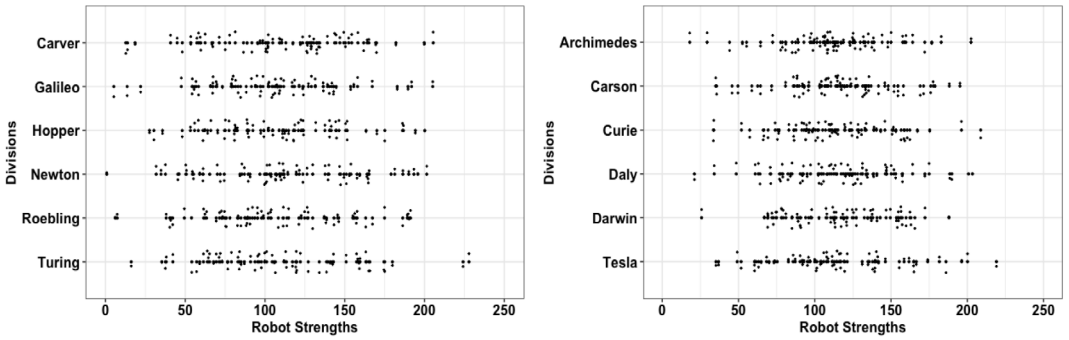

Fig. 2. Dot plots of estimated robot strengths from the OPR model for the divisions of the 2018 FRC Houston and Detroit championships with vertical jitter.

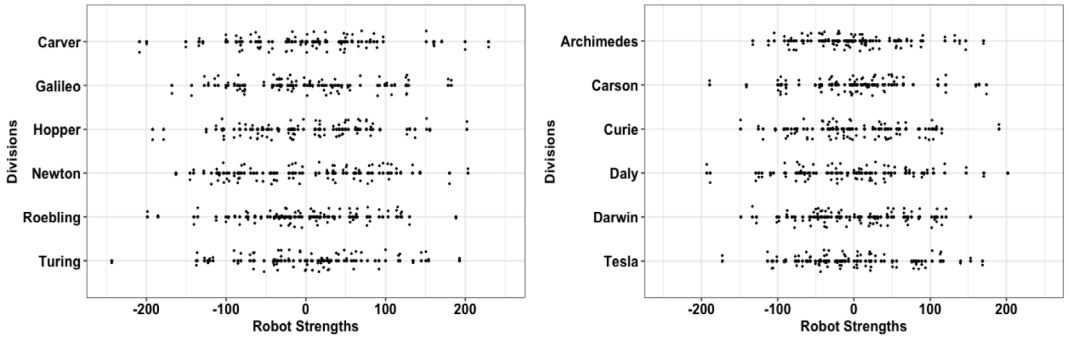

Fig. 3. Dot plots of estimated robot strengths from the WMPR model for the divisions of the 2018 FRC Houston and Detroit championships with vertical jitter.

Remark 4. Under the assumption that chess players of comparable skill levels play against each other, Sismanis (2010) developed the Elo++ rating system to avoid overfitting the ratings of players in the Elo rating system. Since robots are randomly assigned to matches and alliances in the qualification stage, there is no explicit neighbor for each robot and, hence, this technique is inappropriate for estimating the OPRC and WMPRC models. Furthermore, it is rather subjective for the defined neighbor averages in the introduced ℓ_2–regularization. □

3.2. Predictive ability

In this subsection, we consider all possible regression models of the form

$$Y = X^c \beta^c + \varepsilon, \tag{14}$$

where $X^c = (X^c_{(1)}, \ldots, X^c_{(c)})$ and $\beta^c = (\beta^c_1, \ldots, \beta^c_c)^\top$ with $X^c_{(j)}$ being defined as $X^{c_0}_{(j)}$ based on $X_{(i)}$'s, $c = 1, \ldots, K$. As in many scientific studies, sensitivity and specificity are commonly used measures of diagnostic accuracy. We further employ these two measures to detect the condition when the blue alliance defeats the red alliance and the condition when the red alliance defeats the blue alliance, respectively. Since robots are randomly assigned to one of two alliances, the expected value of the win indicator D_0 should be 0.5. It follows that the accuracy, i.e. the weighted average of sensitivity and specificity, is a reasonable measure to assess the predictive ability of a model. More precisely, such a measure is the expected proportion of correct predictions. Given the number of clusters c, clusters of robots $\widehat{g}$, and an estimator $\widehat{\beta}^c$ of β^c, the accuracy of a predictor based on model (14) is defined as

$$\mathrm{AC}(c) = P\big(\mathrm{sign}\big(Y^{\mathrm{R}}_0 - Y^{\mathrm{B}}_0\big) \cdot \mathrm{sign}\big(\widehat{P}^c\big(x^{\mathrm{B}}_0, x^{\mathrm{R}}_0\big) - 0.5\big) > 0\big) + 0.5P\big(\mathrm{sign}\big(Y^{\mathrm{R}}_0 - Y^{\mathrm{B}}_0\big) \cdot \mathrm{sign}\big(\widehat{P}^c\big(x^{\mathrm{B}}_0, x^{\mathrm{R}}_0\big) - 0.5\big) = 0\big), \tag{15}$$

where $\widehat{P}^c(x^{\mathrm{B}}_0, x^{\mathrm{R}}_0) = 1 - \widehat{F}^c\big(-(x^{\mathrm{R}c}_0 - x^{\mathrm{B}c}_0)^\top \widehat{\beta}^c\big)$ with $\widehat{F}^c(v) = \sum_{s=1}^M I(e^c_s \le v)/m$ and $e^c_s = (Y^{\mathrm{R}}_s - Y^{\mathrm{B}}_s) - (X^{\mathrm{R}c}_s - X^{\mathrm{B}c}_s)^\top \widehat{\beta}^{(c)}$, $c = 1, \ldots, K$, $s = 1, \ldots, M$. We note that the second term on the right-hand side of (15) is used to deal with the problem of ties. Provided that c_0 and g are known, the empirical distribution function $\widehat{F}^{c_0}(v)$ of residuals is shown by Mammen (1996) to converge uniformly to the distribution function of errors under some suitable conditions. For the difference in score between the red and blue alliances, i.e. $(Y^{\mathrm{R}}_0 - Y^{\mathrm{B}}_0)$, the following mean square prediction error is adopted:

$$\mathrm{MSPE}(c) = E\big[\big(\big(Y^{\mathrm{R}}_0 - Y^{\mathrm{B}}_0\big) - \big(x^{\mathrm{R}c}_0 - x^{\mathrm{B}c}_0\big)^\top \widehat{\beta}^c\big)^2\big], \ c = 1, \ldots, K. \tag{16}$$

As in the context of regression analysis, this measure is found to be useful in exploring the effect of potential influential observations and in selecting several competing models.

In the qualification stage, AC(c) and MSPE(c) are estimated by the leave-one-match-out cross-validation estimates

$$\widehat{\mathrm{AC}}(c)=\frac{1}{M}\sum_{s=1}^{M}D_s(c) \text{ and } \widehat{\mathrm{MSPE}}(c)=\frac{1}{M}\sum_{s=1}^{M}\big((Y_s^{\mathrm{R}}-Y_s^{\mathrm{B}})-(X_s^{\mathrm{R}c}-X_s^{\mathrm{B}c})^{\top}\widehat{\beta}^{c}_{-s}\big)^2, \text{ respectively, } c=1,\ldots,K, \tag{17}$$

where $D_s(c)=I\big(\mathrm{sign}(Y_s^{\mathrm{R}}-Y_s^{\mathrm{B}})\cdot\mathrm{sign}(\widehat{P}^{c}_{-s}(X_s^{\mathrm{B}},X_s^{\mathrm{R}})-0.5)>0\big)+0.5I\big(\mathrm{sign}(Y_s^{\mathrm{R}}-Y_s^{\mathrm{B}})\cdot\mathrm{sign}(\widehat{P}^{c}_{-s}(X_s^{\mathrm{B}},X_s^{\mathrm{R}})-0.5)=0\big)$, $\widehat{P}^{c}_{-s}(X_s^{\mathrm{B}},X_s^{\mathrm{R}})=1-\widehat{F}^{c}_{-s}\big(-(X_s^{\mathrm{R}c}-X_s^{\mathrm{B}c})^{\top}\widehat{\beta}^{c}_{-s}\big)$, and $\big(\widehat{\beta}^{c}_{-s},\widehat{F}^{c}_{-s}(v)\big)$ is computed as $\big(\widehat{\beta}^{c},\widehat{F}^{c}(v)\big)$ with match s being removed, $s=1,\ldots,M$. Instead of using data $\{Y_{-s},X^{c}_{-s}\}$, $\big(\widehat{\beta}^{c}_{-s},\widehat{F}^{c}_{-s}(v)\big)$ can be directly obtained through the relation between $\widehat{\beta}^{c}_{-s}$ and $\{\widehat{\beta}^{c},X^{c},e^{c}_{s}\}$ for each s. The computational details are relegated to Appendix. Let AC_1 and AC_2 stand for the accuracies of any two generic models of a division with the corresponding estimates $\widehat{\mathrm{AC}}_1=\sum_{s=1}^{M}D_{1s}/M$ and $\widehat{\mathrm{AC}}_2=\sum_{s=1}^{M}D_{2s}/M$, where D_{1s}'s and D_{2s}'s are defined as $D_s(c)$'s. It is ensured by Theorem 1 in Chiang and Chiu (2012) that $\sqrt{M}(\widehat{\mathrm{AC}}_{\ell}-\mathrm{AC}_{\ell})/\widehat{\sigma}_{\ell}$ can be approximated by the standard normal distribution, where $\widehat{\sigma}^2_{\ell}=\sum_{s=1}^{M}(D_{\ell s}-\widehat{\mathrm{AC}}_{\ell})^2/M$, $\ell=1,2$. Using this property, an approximate $(1-\alpha)$-confidence interval for AC_{ℓ} is, thus, constructed as follows:

$$\widehat{\mathrm{AC}}_{\ell}\pm z_{1-\alpha/2}\frac{\widehat{\sigma}_{\ell}}{\sqrt{M}},\ \ \ell=1,2, \tag{18}$$

where z_q is the qth quantile value of the standard normal distribution. For the hypotheses $H_0:\mathrm{AC}_1=\mathrm{AC}_2$ versus $H_A:\mathrm{AC}_1\neq\mathrm{AC}_2$ (or the hypotheses $H_0:\mathrm{AC}_1\leq\mathrm{AC}_2$ versus $H_A:\mathrm{AC}_1>\mathrm{AC}_2$), the following test statistic is proposed:

$$T=\frac{\sqrt{M}\big(\widehat{\mathrm{AC}}_1-\widehat{\mathrm{AC}}_2\big)}{\widehat{\sigma}_d}, \text{ where } \widehat{\sigma}^2_d=\frac{1}{M}\sum_{s=1}^{M}\big((D_{1s}-D_{2s})-(\widehat{\mathrm{AC}}_1-\widehat{\mathrm{AC}}_2)\big)^2. \tag{19}$$

In our test, H_0 is rejected with an approximate significance level α whenever $|T|>Z_{1-\alpha/2}$ (or $T>Z_{1-\alpha}$).

In the playoff stage, we let K^* and M^* stand for the number of robots and the number of matches, respectively, in a division; and $Y^{*(t)}=\big(Y_1^{*(t)},\ldots,Y_{M^*}^{*(t)}\big)^{\top}$ and $X^{*(t)}=\big(X_1^{*(t)},\ldots,X_{M^*}^{*(t)}\big)^{\top}$ the designed response vector and covariate matrix with (t) being either B or R. By treating qualification data as training data, the measures AC(c) and MSPE(c) of an estimation procedure on playoff data are defined as

$$\mathrm{AC}(c)=\frac{1}{M^*}\sum_{s=1}^{M^*}D_s^*(c) \text{ and } \mathrm{MSPE}(c)=\frac{1}{M^*}\sum_{s=1}^{M^*}\big((Y_s^{*\mathrm{R}}-Y_s^{*\mathrm{B}})-(X_s^{*\mathrm{R}c}-X_s^{*\mathrm{B}c})^{\top}\widehat{\beta}^{c}\big)^2, c=1,\ldots,K^*, \tag{20}$$

where $D_s^*(c)=I\big(\mathrm{sign}(Y_s^{*\mathrm{R}}-Y_s^{*\mathrm{B}})\cdot\mathrm{sign}(\widehat{P}^{c}(X_s^{*\mathrm{B}c},X_s^{*\mathrm{R}c})-0.5)>0\big)+0.5I\big(\mathrm{sign}(Y_s^{*\mathrm{R}}-Y_s^{*\mathrm{B}})\cdot\mathrm{sign}(\widehat{P}^{c}(X_s^{*\mathrm{B}c},X_s^{*\mathrm{R}c})-0.5)=0\big)$ with $(X_s^{*\mathrm{B}c},X_s^{*\mathrm{R}c})$'s being computed as $(X_s^{\mathrm{B}c},X_s^{\mathrm{R}c})$'s. As shown in our application, the predictive ability of the current and proposed models is poor for match outcomes in the playoff stage. Due to non-random assignment of robots to matches, some confounders cannot be treated merely as nuisance variables in model fitting. This partly explains poor capacity in prediction.

3.3. Estimation procedures

For the unknown parameters c_0, g, and β^{c_0} in model (12), it is usually impractical to fit all possible regression models in (14). The total number of these model candidates can be further shown to be

$$S_K=\sum_{c=1}^{K}\frac{1}{c!}\sum_{j=0}^{c-1}(-1)^j\binom{c}{j}(c-j)^K, \tag{21}$$

which is the Stirling number of the second kind as shown by Marx (1962) and Salmeri (1962). In the 2018 FRC Houston and Detroit championships, S_K is about 1.67×10^{69} for $K=67$ and 3.66×10^{70} for $K=68$. In this study, two effective estimation procedures are developed to avoid the computational complexity and cost associated with the selection of an appropriate model from a huge number of possible model candidates indexed by the combinations of variant numbers of clusters and clusters of robots.

We first propose the following estimation procedure (Method 1) for model (12):

Step 1. Fit a regression model in (8) (e.g. the OPR and WMPR models) and compute the LSE $\widehat{\beta}$ of β.
Step 2. Perform the centroid linkage clustering method of Sokal and Michener (1958) on $\widehat{\beta}$ and obtain cluster estimators $\widehat{g}=(\widehat{g}_1,\ldots,\widehat{g}_K)^{\top}$ with $\widehat{g}_i$'s $\in\{1,\ldots,c\}$.
Step 3. Fit a regression model $Y=X^{c}\beta^{c}+\varepsilon$ and compute the LSE $\widehat{\beta}^{c}$ of β^{c}, $c=1,\ldots,K$.
Step 4. Compute an estimate $\widehat{\mathrm{AC}}(c)$ of AC(c) (or an estimate $\widehat{\mathrm{MSPE}}(c)$ of MSPE(c)) and derive the maximizer $\widehat{c}=\arg\max_c\widehat{\mathrm{AC}}(c)$ (or the minimizer $\widehat{c}^*=\arg\min_c\widehat{\mathrm{MSPE}}(c)$).
Step 5. Estimate (c_0,g,β^{c_0}) by $(\widehat{c},\widehat{g},\widehat{\beta}^{\widehat{c}})$.

The reason of using $\widehat{\beta}$ as an initial estimator is mainly based on the validity of model (8), which is an overparameterized version of the proposed model for $K > c_0$, and the consistency of $\widehat{\beta}$ to β. As shown by Portnoy (1984), its convergence rate is the square root of the ratio of the number of robots to the number of observations. Owing to the constant strength of robots in the same cluster, i.e. β_i's $\in \{\beta_1^{c_0}, \ldots, \beta_{c_0}^{c_0}\}$, the centroid linkage clustering method, which is one of the hierarchical clustering methods (Gordon, 1987), is reasonably used to measure the distance between any two clusters. For $c \geq c_0$, it is further ensured that $\widehat{\beta}^c$, which is a consistent estimator of β^c with β_i^c's $\in \{\beta_1^{c_0}, \ldots, \beta_{c_0}^{c_0}\}$, will be more and more precise as c decreases to c_0. Thus, the determination of c_0 can be naturally transferred to the model selection problem in such a dimension reduction of the parameter space. In the above estimation procedure, the LSE of β_0^1 is directly derived as $\widehat{\beta}^1 = \sum_{s=1}^{M}(Y_s^{\mathrm{R}} + Y_s^{\mathrm{B}})/6M$ for the OPRC model and $\widehat{\beta}^1 = 0$ for the WMPRC model. Different from existing criteria (e.g. Krzanowski and Lai, 1985; Tibshirani et al., 2001; Sugar and James, 2003; Wang, 2010) in the cluster analysis, the optimal number of clusters c_0 is determined by either maximizing the leave-one-match-out cross-validation estimate of the accuracy $\mathrm{AC}(c)$ or minimizing the leave-one-match-out cross-validation estimate of the mean square prediction error $\mathrm{MSPE}(c)$ with respect to the number of clusters c, where the left-out match (testing data) plays as the role of a future run and the remaining matches (training data) are used to build up competing models. Although the properties $\mathrm{AC}(c_0) > \mathrm{AC}(c) + O(M^{-1})$ and $\mathrm{MSPE}(c_0) < \mathrm{MSPE}(c) + O(M^{-1})$ hold for $c < c_0 \ll K$, the cross-validation criterion, which is akin to the Akaike information criterion (AIC) (Akaike, 1974), is inconsistent in model selection because of the properties $\mathrm{AC}(c) = \mathrm{AC}(c_0) + O(M^{-1})$ and $\mathrm{MSPE}(c) = \mathrm{MSPE}(c) + O(M^{-1})$ for $c_0 < c \ll K$. However, in our application, the imprecise estimates of robot strengths in a set of nested models of the form (14) are found to result in $\widehat{\mathrm{AC}}(c_1) > \widehat{\mathrm{AC}}(c_2)$ for $c_2 > c_1 \geq \widehat{c}$ and $\widehat{\mathrm{MSPE}}(c_1) < \widehat{\mathrm{MSPE}}(c_2)$ for $c_2 > c_1 \geq \widehat{c}^*$. This implicitly implies that $\widehat{c}$ and $\widehat{c}^*$ should be much smaller than K and close to c_0. A more thorough study warrants future research. It is notable that $\widehat{\mathrm{AC}}(c)$ and $\widehat{\mathrm{MSPE}}(c)$ are not only used to assess the predictive ability of models, but also to estimate the number of clusters c_0 in model (12).

In fact, clusters of robots $\widehat{g}_i$'s can be obtained by performing the cluster analysis on the estimator $\widetilde{\beta}^{(c+1)}$ sequentially with $c = K-1, \ldots, 2$. As an alternative of Method 1, the second estimation procedure (Method 2) is further proposed as follows:

Step 1. Fit a regression model $Y = X\beta + \varepsilon$ and compute the LSE $\widehat{\beta}$ of β and $\widehat{\mathrm{AC}}(K)$ (or $\widehat{\mathrm{MSPE}}(K)$).
Step 2. Perform the centroid linkage clustering method on $\widetilde{\beta}^K \stackrel{\triangle}{=} \widehat{\beta}$ and obtain cluster estimators $\widehat{g} = (\widehat{g}_1, \ldots, \widehat{g}_K)^\top$ with $\widehat{g}_i$'s $\in \{1, \ldots, K-1\}$.
Step 3. Fit a regression model $Y = X^{K-1}\beta^{K-1} + \varepsilon$ and compute the LSE $\widehat{\beta}^{K-1}$ of β^{K-1} and $\widehat{\mathrm{AC}}(K-1)$ (or $\widehat{\mathrm{MSPE}}(K-1)$).
Step 4. Repeat Steps 2–3 for $\widehat{g}$, $\widetilde{\beta}^c$, $\widehat{\beta}^{c-1}$, and $\widehat{\mathrm{AC}}(c-1)$ (or $\widehat{\mathrm{MSPE}}(c-1)$), $c = K-1, \ldots, 2$.
Step 5. Estimate c_0, g, and β^{c_0} by $\widehat{c} = \arg\max_c \widehat{\mathrm{AC}}(c)$ (or $\widehat{c}^* = \arg\min_c \widehat{\mathrm{MSPE}}(c)$), $\widehat{g}$, and $\widehat{\beta}^{\widehat{c}}$, respectively.

As we can see, clusters of robots $\widehat{g}_i$'s are obtained by performing cluster analysis on $\widetilde{\beta}^K$ in Method 1 but on $\widetilde{\beta}^{c+1}$, $c = K-1, \ldots, 2$, in Method 2. It is notable that Method 2 shares the asymptotic properties of Method 1 because this estimation procedure uses the same consistent initial estimator. Although both estimation procedures might produce different estimated numbers of clusters, their estimated accuracies for the proposed model are close to each other in our application.

In the proposed model (12), the multiple comparison procedures such as those proposed by Hochberg and Tamhane (1987), Hsu (1996), and Hothorn et al. (2008), are possible avenues for the determination of the number of clusters and clusters of robots. However, how to control the overall type I error rate for the simultaneous inference of $\{\beta_i - \beta_j : i \neq j\}$ is still a challenging task. Furthermore, such an inference procedure might not be able to achieve the prediction purpose. As indicated in the introduction, a small ratio of the number of observations to the number of robots will affect the precision of the LSE $\widehat{\beta}$ of β in (8). It tends to choose a small number of clusters and, hence, lead to poor prediction on match outcomes.

Remark 5. The formulation in (14) for the WMPRC model can be rewritten as $Y = \bar{X}^c\beta^{*c} + \varepsilon$, where $\bar{X}^c$ is defined as $\bar{X}$ with the jth column being $X_{(j)}^c - (n_1/n_c)X_{(c)}^c$, and $\beta^{*c} = (\beta_1^c, \ldots, \beta_{c-1}^c)^\top$ with $\beta_c^c = -\sum_{j=1}^{c-1} n_j\beta_j^c/n_c$ and $n_j = \sum_{i=1}^{n} I(g_i = j), j = 1, \ldots, c$. As in the context of regression analysis, the LSE of β^c in the OPRC model is

$$\widehat{\beta}^c = \left(X^{c\top}X^c\right)^{-1}X^{c\top}Y \tag{22}$$

and the LSE of β^c in the WMPRC model is

$$\widehat{\beta}^c = \begin{pmatrix} \widehat{\beta}^{*c} \\ \widehat{\beta}_c^c \end{pmatrix} \text{ with } \widehat{\beta}^{*c} = \left(\bar{X}^{c\top}\bar{X}^c\right)^{-1}\bar{X}^{c\top}Y \text{ and } \widehat{\beta}_c^c = \frac{-\sum_{j=1}^{c-1} n_j\widehat{\beta}_j^{*c}}{n_c} \text{ for } c \geq 2. \quad \square \tag{23}$$

Remark 6. Although the Bayesian information criterion (BIC) (Schwarz, 1978) has been widely used in model selection, Giraud (2015) showed that the BIC is valid only when M is much larger than K, which is not the case for this scenario. Furthermore, the BIC is infeasible in our setup with the lack of a particular distribution assumption on Y_s and D_s, $s = 1, \ldots, M$. As we can see, existing methods in the latent cluster analysis (Lazarsfeld and Henry, 1968; Goodman, 1974; Collins and Lanza, 2010) are inapplicable to the proposed model because of their focus mainly on identifying the

underlying clusters of Y_s's rather than those of β_i's. Furthermore, when c is greater than 1, there exists a huge number of possible class membership combinations $\left(\sum_{j=0}^{c-1}(-1)^j\binom{c}{j}(c-j)^K\right)/c!$ for g_i's in the perspective of latent class analysis. For different combinations of g_i's, it also needs to carry out a complicated computational task in deriving the corresponding design covariates $X_{(j)}^c$'s and estimates $\widehat{\beta}_j^c$'s. □

3.4. Agreement assessment

Let R_0 be the collection of FRC ratings of all, playoff, or FRC top-8 robots; K_0 the size of R_0; and $\widetilde{\beta}_0$ the estimated robot strengths of robots in R_0. To assess the agreement between FRC ratings and model-based robot strengths, the rank correlation, which was first proposed by Han (1987), of R_0 and $\widetilde{\beta}_0$ is computed as

$$\mathrm{RC}(R_0, \widetilde{\beta}_0) = \frac{1}{K_0(K_0-1)} \sum\sum_{i\neq j} I\big(\mathrm{sign}(R_i - R_j)\cdot \mathrm{sign}(\widetilde{\beta}_i - \widetilde{\beta}_j) > 0\big) + 0.5I\big(\mathrm{sign}(R_i - R_j)\cdot \mathrm{sign}(\widetilde{\beta}_i - \widetilde{\beta}_j) = 0\big), \tag{24}$$

where R_i's and $\widetilde{\beta}_i$'s are the corresponding elements of R_0 and $\widetilde{\beta}_0$. In fact, this rank-based measure is particularly useful to investigate the monotonic association of two ordinal scale measurements. Other measures such as Kendall's τ (Kendall, 1938) and spearman's ρ (Spearman, 1904) can also be used to assess the agreement between R_0 and $\widetilde{\beta}_0$.

As in the contexts of pattern recognition and information retrieval, we assess the agreement between FRC top-8 robots and model-based top-N robots, which are rated by estimated robot strengths, by the precision and recall measures from Perry et al. (1955). Let RS_0 and $\mathrm{RS}(N)$ stand for the corresponding sets of FRC top-8 robots and model-based top-N robots. The precision and recall measures of RS_0 and $RS(N)$ are further given by

$$\mathrm{Pr}(\mathrm{RS}_0, \mathrm{RS}(N)) = \frac{|\mathrm{RS}_0 \cap \mathrm{RS}(N)|}{N} \text{ and } \mathrm{Re}(\mathrm{RS}_0, \mathrm{RS}(N)) = \frac{|\mathrm{RS}_0 \cap \mathrm{RS}(N)|}{8}, \text{ respectively.} \tag{25}$$

In diagnostic tests, the precision and recall are termed as the positive predictive value by Vecchio (1966) and sensitivity by Yerushalmy (1947), respectively. It is notable that the sensitivity is a measure of the intrinsic accuracy of a test whereas the positive predictive value is only a function of the accuracy and prevalence.

3.5. Stability assessment

In this study, we assess how many matches are needed in the qualification stage to produce a clear picture about robot strengths. This is mainly to make a recommendation on an optimal number of matches in future tournaments to improve planning and logistics. Let $Y_{[\ell]}^{(t)}$ and $X_{[\ell]}^{(t)}$ be the corresponding vector and matrix formed by the first M_ℓ elements of $Y^{(t)}$ and the first M_ℓ rows of $X^{(t)}$ with (t) being either B or R and $M_\ell = \lceil \frac{\ell\times K}{6} \rceil$, $\ell = 6, \ldots, m_0$. Based on data $\left\{Y_{[\ell]}^{\mathrm{B}}, Y_{[\ell]}^{\mathrm{R}}, X_{[\ell]}^{\mathrm{B}}, X_{[\ell]}^{\mathrm{R}}\right\}$ and the estimators $(\widehat{c}, \widehat{g})$ (or $(\widehat{c}^*, \widehat{g})$) of (c_0, g), we define $\widetilde{\beta}_{[\ell]}$ as the robot strength estimates or the top-8 robot strength estimates obtained from the LSE of β^{c_0} in model (12), $\ell = 6, \ldots, m_0$.

Same with the formulations of $\widehat{\mathrm{AC}}(c)$ and $\widehat{\mathrm{MSPE}}(c)$ in (17), $\widehat{\mathrm{AC}}_{[\ell]}(\widehat{c})$ and $\widehat{\mathrm{MSPE}}_{[\ell]}(\widehat{c}^*)$ are computed based on data $\left\{Y_{[\ell]}^{\mathrm{B}}, Y_{[\ell]}^{\mathrm{R}}, X_{[\ell]}^{\mathrm{B}}, X_{[\ell]}^{\mathrm{R}}\right\}$ and $\widetilde{\beta}_{[\ell]}$, $\ell = 6, \ldots, m_0$. The stability of estimated robot strengths and accuracies can be assessed through $\{\mathrm{RC}(\widetilde{\beta}_{[\ell]}, \widetilde{\beta}_{[\ell+1]}) : \ell = 6, \ldots, (m_0 - 1)\}$ and $\{\widehat{\mathrm{AC}}_{[\ell]}(\widehat{c})$ (or $\widehat{\mathrm{MSPE}}_{[\ell]}(\widehat{c}^*)) : \ell = 6, \ldots, m_0\}$, respectively, where the rank correlation function $\mathrm{RC}(\cdot,\cdot)$ is defined in (24). Given these measures and a pre-determined tolerance threshold, we can further determine an appropriate number of matches in the qualification stage.

4. An application to the 2018 FRC championships

In this section, the OPR and WMPR models in Sections 2.2 and 2.4–2.5 and the OPRC and WMPRC models in Section 3.1 are applied to the 2018 FRC Houston and Detroit championships. As discussed in Remark 3, the WMPRR model is the WMPR model with random robot strengths. Its flaws are also investigated in this application. As of the 2018 competitions, each championship site is composed of six divisions: Carver, Galileo, Hopper, Newton, Roebling, and Turing in Houston; and Archimedes, Carson, Curie, Daly, Darwin, and Tesla in Detroit. For the determination of the number of clusters and clusters of robots, the developed estimation procedures (Method 1, Method 2) in Section 3.3 are denoted by (OPRC1,OPRC2) for the OPRC model and (WMPRC1,WMPRC2) for the WMPRC model. To avoid verbosity, the corresponding estimation procedures for the OPR and WMPR models are hereinafter denoted by OPR and WMPR estimation procedures. In Table 1, we summarize the number of matches and the number of robots of each division in the qualification stage and the playoff stage. Table 2 further displays the estimated numbers of clusters from the OPRC1, OPRC2, WMPRC1, and WMPRC2 estimation procedures. Apparently, the estimated numbers of clusters are relatively small compared to the number of robots. It is also shown in Figs. 4 and 5 that there exist an inverted U-shaped relationship between the estimated accuracy and the number of clusters and a U-shaped relationship between the estimated mean square prediction error and the number of clusters.

Since the research interest mainly focuses on predicting match outcomes, our investigation is based on the estimated number of clusters $\widehat{c}$. In the twelve divisions, the corresponding ratios of the number of observations to the estimated number of clusters range from 22.33 to 45.33 for the WMPRC model and from 6.48 to 22.67 for the WMPRC model.

Table 1
The number of matches M and the number of robots K in the qualification stage and the number of matches M^* and the number of robots K^* in the playoff stage.

Tournament	Division	M	K	M^*	K^*
Houston	Carver	114	68	17	28
	Galileo	114	68	17	28
	Hopper	114	68	17	28
	Newton	112	67	14	28
	Roebling	112	67	15	28
	Turing	112	67	18	27
Detroit	Archimedes	114	68	16	27
	Carson	114	68	17	27
	Curie	112	67	16	27
	Daly	114	68	16	26
	Darwin	112	67	18	25
	Tesla	112	67	17	27

Table 2
The estimated numbers of clusters $\widehat{c}$ and $\widehat{c}^*$, in which the numbers within the brace are the maximum and minimum cluster sizes, of c_0 from the OPRC1, OPRC2, WMPRC1, and WMPRC2 estimation procedures on qualification data.

Tournament	Division	OPRC1		OPRC2		WMPRC1		WMPRC2	
		$\widehat{c}$	$\widehat{c}^*$	$\widehat{c}$	$\widehat{c}^*$	$\widehat{c}$	$\widehat{c}^*$	$\widehat{c}$	$\widehat{c}^*$
Houston	Carver	3(46, 3)	11(15, 1)	10(18, 1)	11(15, 1)	10(21, 1)	11(14, 1)	12(12, 1)	11(15, 1)
	Galileo	12(19, 1)	9(19, 1)	10(18, 1)	8(18, 1)	7(22, 1)	8(17, 1)	8(17, 1)	8(17, 1)
	Hopper	7(23, 1)	9(15, 1)	7(25, 1)	9(15, 1)	8(20, 1)	10(12, 1)	21(9, 1)	10(12, 1)
	Newton	7(23, 1)	10(17, 1)	8(17, 1)	9(17, 1)	14(10, 1)	13(10, 1)	14(9, 1)	10(17, 1)
	Roebling	7(23, 2)	9(15, 1)	7(18, 2)	8(18, 2)	11(17, 1)	10(17, 1)	11(16, 1)	11(16, 1)
	Turing	8(16, 1)	8(16, 1)	11(13, 1)	9(14, 1)	11(20, 1)	12(12, 1)	10(25, 1)	8(14, 1)
Detroit	Archimedes	15(20, 1)	8(26, 1)	14(20, 1)	8(20, 1)	14(15, 1)	13(15, 1)	9(26, 2)	6(16, 2)
	Carson	4(35, 8)	10(13, 2)	11(16, 2)	9(16, 2)	10(23, 1)	12(12, 1)	12(16, 1)	10(22, 1)
	Curie	16(15, 1)	12(15, 1)	16(13, 1)	10(17, 1)	8(23, 1)	12(12, 1)	6(25, 1)	10(13, 1)
	Daly	6(31, 1)	9(19, 1)	6(31, 2)	8(19, 2)	14(14, 1)	11(14, 1)	16(9, 1)	11(14, 1)
	Darwin	9(13, 1)	7(19, 1)	8(16, 1)	9(16, 1)	8(15, 1)	9(15, 1)	8(21, 1)	8(21, 1)
	Tesla	8(30, 1)	9(20, 1)	6(30, 1)	8(15, 1)	7(22, 1)	12(11, 1)	7(25, 1)	12(15, 1)

Table 3 displays the estimated accuracies of the form in (17) from different estimation procedures on qualification data for the divisions of the Houston and Detroit tournaments. For the accuracies of the investigated models, approximate 0.95-confidence intervals, which are constructed as in (18), are further provided in Table 3. Although there is no significant difference in the accuracies from each estimation procedure for all divisions in the same tournament, it does not mean that the results across divisions can be aggregated to obtain a more robust estimate of the accuracy. This is mainly because the underlying models in some divisions might be completely different. Moreover, the overall estimated accuracies from the AS, OPR, OPRC1, OPRC2, WMPR, WMPRC1, WMPRC2, and WMPRR estimation procedures are computed as 70.3%, 71.0%, 85.0%, 85.2%, 69.5%, 89.0%, 89.3%, and 70.1%. Based on the test statistic in (19), it is shown that the accuracies of the OPRC and WMPRC models are significantly higher than those of the OPR and WMPR models in all divisions (p-values ≤ 0.007). As we can see, the accuracies of the AS, OPR, and WMPR models are not significantly different in most of the divisions (p-values ≥ 0.127), except that the accuracy of the WMPR model is significantly lower than the accuracies of the AS model in the Archimedes and Darwin divisions (p-values $= 0.003$ and 0.029) and the OPR model in the Archimedes and Daly divisions (p-values $= 0.007$ and 0.014). The performance of the WMPRR model is only comparable to that of the WMPR model. Its poor prediction is probably caused by over-simplifying the heterogeneous feature of robot strengths. The performance of the OPRC1 and OPRC2 estimation procedures and that of the WMPRC1 and WMPRC2 estimation procedures are also found to be comparable in all divisions (p-values ≥ 0.158). For the divisions of the Houston and Detroit tournaments, the estimated accuracies of the OPRC and WMPRC models are about $8\% - 19\%$ and $14\% - 26\%$ higher than those of the OPR and WMPR models, respectively. Although the predictive ability of the WMPRC model is generally better than that of the OPRC model, their accuracies are not significantly different in all divisions (p-values ≥ 0.079). Despite using a smaller data set, it is possible for the WMPRC model to produce a more meaningful prediction compared to the OPRC model, which does not take into account the feature of paired comparisons. In addition to these findings, we note that the estimated numbers of clusters from the WMPRC1 and WMPRC2 estimation procedures are different in most of the divisions, whereas the predictive capacities of both models are close to each other. The same conclusion can be drawn for the OPRC1 and OPRC2 estimation procedures. As emphasized in Section 3.3, the accuracy and mean square prediction error tend to overestimate the corresponding numbers of clusters in the OPRC and WMPRC models.

Provided that the OPRC (or WMPRC) model is valid for all divisions in the same tournament, we can add up their estimated accuracy functions in (17) to estimate the number of clusters c_0. The estimated numbers of clusters from the OPRC1, OPRC2, WMPRC1, and WMPRC2 estimation procedures are 8, 10, 11, and 15, respectively, for all divisions

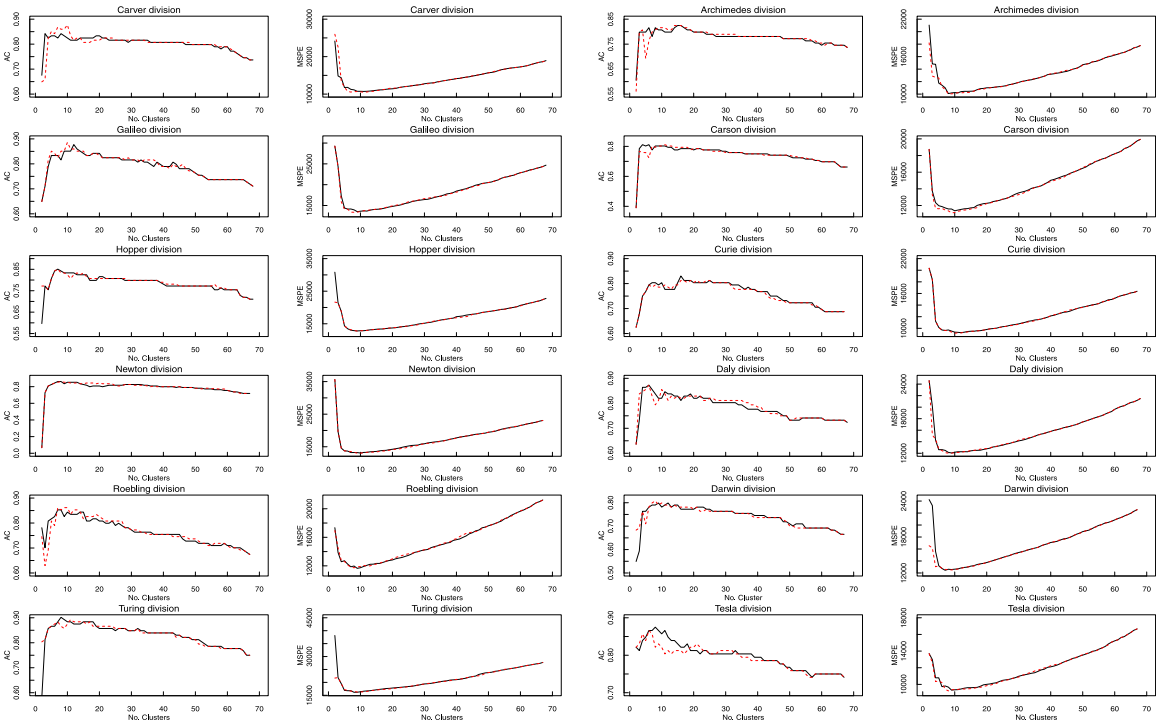

Fig. 4. The estimated accuracy and mean square prediction error curves from the OPRC1 (black-solid line) and OPRC2 (red-dashed line) estimation procedures for the divisions of the 2018 FRC Houston and Detroit championships.

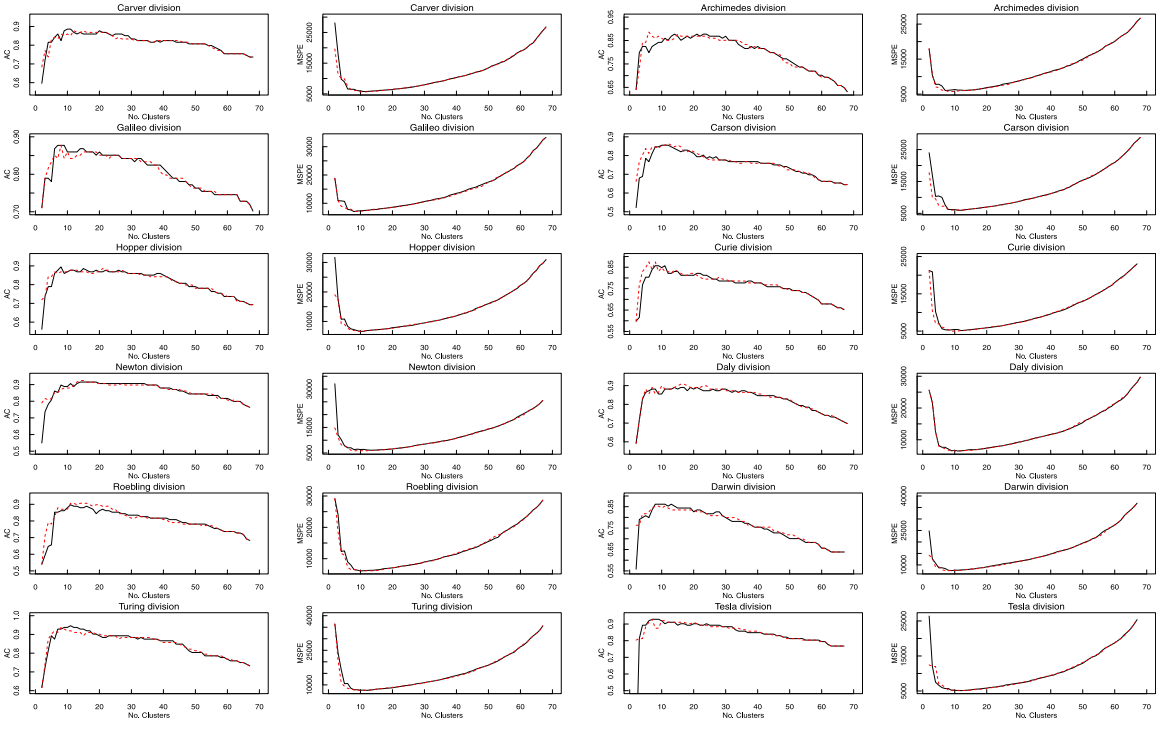

Fig. 5. The estimated accuracy and mean square prediction error curves from the WMPRC1 (black-solid line) and WMPRC2 (red-dashed line) estimation procedures for the divisions of the 2018 FRC Houston and Detroit championships.

Table 3
The estimated accuracies (approximate 0.95-confidence intervals) from the AS, OPR, OPRC1, OPRC2, WMPR, WMPRC1, WMPRC2, and WMPRR estimation procedures on qualification data.

Tournament	Division	AS	OPR	OPRC1	OPRC2	WMPR	WMPRC1	WMPRC2	WMPRR
Houston	Carver	68(58.9,76.2)%	74(65.6,81.8)%	84(77.5,90.9)%	88(81.7,93.8)%	74(65.6,81.8)%	89(82.7,94.5)%	88(81.7,93.8)%	69(60.5,77.5)%
	Galileo	68(58.9,76.2)%	71(62.7,79.4)%	88(81.7,93.8)%	89(82.7,94.5)%	70(61.7,78.6)%	88(81.7,93.8)%	88(81.7,93.8)%	71(62.7,79.3)%
	Hopper	73(64.6,81.0)%	71(62.7,79.4)%	85(78.5,91.7)%	85(78.5,91.7)%	69(60.8,77.8)%	89(83.8,95.1)%	89(82.7,94.5)%	72(63.8,80.2)%
	Newton	78(70.5,85.8)%	72(63.6,80.2)%	86(79.8,92.5)%	86(79.8,92.5)%	76(68.5,84.2)%	92(86.4,96.6)%	92(87.6,97.3)%	74(65.9,82.1)%
	Roebling	67(58.7,76.1)%	67(58.7,76.1)%	85(78.7,91.8)%	86(79.8,92.5)%	68(59.7,76.9)%	90(84.2,95.3)%	91(85.3,96.0)%	71(62.6,79.3)%
	Turing	72(64.0,80.6)%	75(66.9,83.1)%	90(84.6,95.7)%	89(83.5,95.0)%	73(65.0,81.5)%	95(90.5,98.8)%	94(89.2,98.3)%	75(67.0,83.0)%
Detroit	Archimedes	75(66.5,82.6)%	74(65.6,81.8)%	82(75.4,89.5)%	82(75.4,89.5)%	63(54.3,72.1)%	88(81.7,93.8)%	89(82.7,94.5)%	71(62.7,79.3)%
	Carson	70(61.3,78.2)%	66(57.6,74.9)%	81(74.0,88.3)%	81(74.0,88.3)%	64(55.7,73.3)%	86(79.1,92.0)%	86(80.1,92.7)%	66(57.3,74.7)%
	Curie	66(57.4,74.8)%	69(60.2,77.3)%	83(76.2,89.9)%	81(74.1,88.4)%	65(56.4,74.0)%	86(79.3,92.1)%	88(81.5,93.5)%	67(58.3,75.7)%
	Daly	64(55.7,73.3)%	72(64.2,80.6)%	87(81.2,93.4)%	87(81.2,93.4)%	70(61.3,78.2)%	89(83.3,94.7)%	91(85.5,96.1)%	68(59.4,76.6)%
	Darwin	71(62.6,79.4)%	67(57.8,75.3)%	80(72.5,87.3)%	81(73.5,88.1)%	64(54.9,72.7)%	86(79.8,92.5)%	85(78.7,91.8)%	66(57.2,74.8)%
	Tesla	72(64.0,80.6)%	74(66.0,82.3)%	88(81.3,93.7)%	87(80.3,92.9)%	77(68.9,84.6)%	93(88.1,97.6)%	93(88.1,97.6)%	71(62.6,79.4)%

Table 4
The estimated accuracies (approximate 0.95-confidence intervals) and accuracies from the OPRC1, OPRC2, WMPRC1, and WMPRC2, estimation procedures, in which a common estimated number of clusters is used for all divisions in the same tournament, on qualification and playoff data, respectively..

Tournament	Division	Qualification data				Playoff data			
		OPRC1	OPRC2	WMPRC1	WMPRC2	OPRC1	OPRC2	WMPRC1	WMPRC2
Houston	Carver	84(77.5,90.9)%	88(81.7,93.8)%	89(82.7,94.5)%	88(81.7,93.8)%	65%	65%	59%	59%
	Galileo	82(74.4,88.7)%	89(82.7,94.5)%	86(79.6,92.4)%	86(79.6,92.4)%	35%	41%	59%	71%
	Hopper	84(77.5,90.9)%	82(75.4,89.5)%	88(81.7,93.8)%	87(80.6,93.1)%	81%	81%	75%	75%
	Newton	86(79.8,92.5)%	86(79.8,92.5)%	91(85.3,96.0)%	92(87.6,97.3)%	57%	71%	71%	71%
	Roebling	85(78.7,91.8)%	86(79.8,92.5)%	90(84.2,95.3)%	91(85.3,96.0)%	57%	60%	40%	40%
	Turing	90(84.6,95.7)%	88(81.3,93.7)%	95(90.5,98.8)%	89(83.5,95.0)%	44%	44%	33%	33%
Detroit	Archimedes	82(74.4,88.7)%	82(74.4,88.7)%	86(79.6,92.4)%	86(79.6,92.4)%	50%	50%	56%	63%
	Carson	81(74.0,88.3)%	80(73.0,87.6)%	86(79.1,92.0)%	86(79.1,92.0)%	76%	76%	76%	71%
	Curie	79(72.1,86.9)%	79(71.0,86.1)%	86(79.3,92.1)%	85(78.3,91.4)%	75%	75%	63%	69%
	Daly	87(81.2,93.4)%	86(79.1,92.0)%	88(82.3,94.1)%	90(84.4,95.4)%	75%	75%	56%	56%
	Darwin	78(70.5,85.8)%	80(72.5,87.3)%	86(79.8,92.5)%	84(77.7,91.1)%	50%	50%	28%	28%
	Tesla	87(80.3,92.9)%	82(75.0,89.3)%	90(84.6,95.7)%	92(86.9,97.0)%	59%	71%	35%	24%

in Houston and 6, 10, 11, and 10, respectively, for all divisions in Detroit. Table 4 further displays the estimated accuracies (approximate 0.95-confidence intervals) and accuracies from these estimation procedures, in which a common estimated number of clusters, say $\bar{c}$, is used for all divisions in the same tournament, on qualification and playoff data, respectively. Our test shows that there is no significant difference between the performance of the estimation procedures with estimated numbers of clusters $\widehat{c}$ and $\bar{c}$ in most of the divisions (p-values $\geq$ 0.079). However, the OPRC1, OPRC2, and WMPRC2 estimation procedures with estimated numbers of clusters $\widehat{c}$ significantly outperform the corresponding estimation procedures with estimated numbers of clusters $\bar{c}$ in the Galileo and Curie divisions (p-values = 0.004 and 0.051), the Curie and Tesla divisions (p-values = 0.041 and 0.047), and the Turing and Archimedes divisions (p-value = 0.012 and 0.041). Based on these results, there is no strong evidence to use the same number of clusters for different divisions of a tournament.

Since robots are not randomly assigned to matches and alliances in the playoff stage, the effects of some confounders might be ignored in model fitting for qualification data. By treating qualification and playoff data as training and testing data, the accuracy of a test on match outcomes in the playoff stage is computed as in (20) and is expected to be low (see Table 5). It can be seen that the predictive ability of the OPR model is comparable with or even better than that of the OPRC model. Except in the Turing, Darwin, and Tesla divisions, the WMPR model is comparable with or better than the WMPRC model. The WMPRR model are further shown to have very poor prediction for match outcomes in the playoff stage. Moreover, the OPR model performs similarly to the AS model except in the Roebling division and outperforms the WMPR model except in the Galileo, Newton, and Archimedes divisions. In Table 4, we also find the poor predictive capacities of the OPRC and WMPRC models with a common number of clusters for all divisions in the same tournament. For all, playoff, and FRC top-8 robots, the corresponding rank correlations of the form in (24) between FRC ratings and estimated robot strengths from the OPR model are generally higher than those between FRC ratings and estimated robot strengths from the rest models (Tables 6–8). This particularly explains why the predictive ability of the OPR model is better than those of the OPRC, WMPR, and WMPRC models in the playoff stage.

For the agreement between FRC ratings and model-based robot strengths of all robots in Table 6 or playoff robots in Table 7, the rank correlations between FRC ratings and robot strengths estimated by the OPR, OPRC1, and OPRC2 estimation procedures are comparable in all divisions. The same conclusion can be further made for the rank correlations between FRC ratings and robot strengths estimated by the WMPR, WMPRC1, and WMPRC2 estimation procedures. Moreover, we find that the rank correlations between FRC ratings and robot strengths estimated by the OPR, OPRC1, and OPRC2 estimation procedures are higher than those between FRC ratings and robot strengths estimated by the WMPR, WMPRC1, and WMPRC2 estimation procedures. Generally speaking, there exist very strong monotonic associations among

Table 5
The accuracies from the AS, OPR, OPRC1, OPRC2, WMPR, WMPRC1, WMPRC2, and WMPRR estimation procedures on playoff data.

Tournament	Division	AS	OPR	OPRC1	OPRC2	WMPR	WMPRC1	WMPRC2	WMPRR
Houston	Carver	71%	65%	65%	65%	59%	59%	59%	60%
	Galileo	47%	41%	41%	41%	71%	59%	59%	57%
	Hopper	75%	81%	81%	81%	75%	75%	75%	70%
	Newton	71%	71%	57%	57%	79%	71%	71%	72%
	Roebling	80%	60%	67%	60%	40%	40%	40%	40%
	Turing	44%	47%	44%	50%	33%	33%	33%	47%
Detroit	Archimedes	63%	50%	50%	50%	63%	59%	63%	62%
	Carson	76%	76%	76%	76%	71%	71%	71%	58%
	Curie	63%	69%	69%	72%	69%	63%	63%	60%
	Daly	63%	75%	75%	75%	56%	56%	56%	59%
	Darwin	44%	50%	50%	50%	28%	33%	28%	34%
	Tesla	53%	65%	71%	71%	29%	47%	41%	43%

Table 6
The rank correlations between FRC ratings and model-based robot strengths, which are estimated by the OPR, OPRC1, OPRC2, WMPR, WMPRC1, and WMPRC2 estimation procedures, of all robots.

Tournament	Division	OPR	OPRC1	OPRC2	WMPR	WMPRC1	WMPRC2
Houston	Carver	77%	68%	77%	69%	68%	69%
	Galileo	82%	81%	81%	78%	77%	77%
	Hopper	77%	76%	77%	76%	75%	76%
	Newton	80%	78%	78%	73%	73%	73%
	Roebling	80%	78%	78%	74%	74%	73%
	Turing	79%	79%	78%	74%	74%	73%
Detroit	Archimedes	82%	81%	81%	77%	76%	75%
	Carson	82%	77%	81%	73%	71%	72%
	Curie	76%	75%	75%	70%	68%	68%
	Daly	79%	76%	77%	73%	73%	73%
	Darwin	78%	77%	76%	72%	72%	72%
	Tesla	79%	76%	75%	75%	74%	74%

Table 7
The rank correlations between FRC ratings and model-based robot strengths, which are estimated by the OPR, OPRC1, OPRC2, WMPR, WMPRC1, and WMPRC2 estimation procedures, of playoff robots.

Tournament	Division	OPR	OPRC1	OPRC2	WMPR	WMPRC1	WMPRC2
Houston	Carver	71%	62%	73%	61%	62%	62%
	Galileo	81%	80%	81%	70%	69%	69%
	Hopper	79%	77%	79%	67%	71%	69%
	Newton	72%	69%	69%	62%	62%	62%
	Roebling	80%	76%	78%	70%	69%	69%
	Turing	70%	73%	72%	68%	67%	68%
Detroit	Archimedes	78%	75%	75%	74%	73%	73%
	Carson	79%	75%	78%	64%	64%	63%
	Curie	74%	74%	74%	66%	67%	67%
	Daly	73%	70%	70%	64%	63%	64%
	Darwin	68%	63%	67%	51%	49%	47%
	Tesla	77%	77%	72%	71%	70%	68%

robot strengths estimated by the OPR, OPRC1, and OPRC2 estimation procedures, and among robot strengths estimated by the WMPR, WMPRC1, and WMPRC2 estimation procedures (Table 9). Due to a very small estimated number of clusters from the OPRC1 estimation procedure in the Carver division, the lower rank correlations are expected between robot strengths estimated by the OPRC1 estimation procedure and robot strengths estimated by the OPR and OPRC2 estimation procedures. As we can see in Table 10, there are relatively lower rank correlations between robot strengths estimated by the OPR, OPRC1, and OPRC2 estimation procedures, and by the WMPR, WMPRC1, and WMPRC2 estimation procedures. For the agreement between FRC ratings and model-based robot strengths of FRC top-8 robots (Table 8), the rank correlation between FRC ratings and robot strengths estimated by the OPR estimation procedure is comparable to or even higher than the rank correlations between FRC ratings and robot strengths estimated by the OPRC1, OPRC2, WMPR, WMPRC1, and WMPRC2 estimation procedures. Overall, the rank correlations between FRC ratings and model-based robot strengths are not strong in most of the divisions.

In Tables 11–12, it is shown that the precisions of the form in (25) between FRC top-8 robots and OPRC1-based top-N robots and between FRC top-8 robots and OPRC2-based top-N robots are comparable to or higher than the precision between FRC top-8 robots and OPR-based top-N robots, $N = 8, 16$. Except for FRC top-8 robots and model-based top-8

Table 8
The rank correlations between FRC ratings and model-based robot strengths, which are estimated by the OPR, OPRC1, OPRC2, WMPR, WMPRC1, and WMPRC2 estimation procedures, of FRC top-8 robots.

Tournament	Division	OPR	OPRC1	OPRC2	WMPR	WMPRC1	WMPRC2
Houston	Carver	75%	66%	79%	61%	68%	63%
	Galileo	50%	54%	54%	39%	46%	48%
	Hopper	71%	66%	66%	57%	66%	63%
	Newton	46%	36%	36%	50%	54%	54%
	Roebling	89%	66%	70%	64%	64%	64%
	Turing	39%	38%	43%	36%	34%	41%
Detroit	Archimedes	61%	64%	64%	57%	61%	57%
	Carson	68%	64%	68%	68%	71%	71%
	Curie	89%	86%	86%	82%	86%	77%
	Daly	82%	77%	77%	82%	82%	82%
	Darwin	64%	55%	63%	54%	50%	50%
	Tesla	50%	48%	48%	57%	66%	52%

Table 9
The rank correlations among robot strengths estimated by the OPR, OPRC1, and OPRC2 estimation procedures, and among robot strengths estimated by the WMPR, WMPRC1, and WMPRC2 estimation procedures.

Tournament	Division	OPR OPRC1	OPR OPRC2	OPRC1 OPRC2	WMPR WMPRC1	WMPR WMPRC2	WMPRC1 WMPRC2
Houston	Carver	73%	92%	73%	92%	95%	92%
	Galileo	93%	93%	93%	90%	92%	90%
	Hopper	90%	89%	89%	91%	97%	91%
	Newton	89%	92%	89%	96%	96%	95%
	Roebling	90%	92%	88%	93%	94%	93%
	Turing	92%	95%	91%	93%	90%	89%
Detroit	Archimedes	94%	93%	93%	95%	88%	87%
	Carson	83%	94%	83%	90%	94%	90%
	Curie	95%	96%	95%	89%	87%	86%
	Daly	85%	86%	85%	94%	96%	94%
	Darwin	93%	92%	90%	92%	91%	90%
	Tesla	87%	85%	85%	89%	88%	86%

Table 10
The rank correlations between robot strengths estimated by the OPR, OPRC1, and OPRC2 estimation procedures and robot strengths estimated by the WMPR, WMPRC1, and WMPRC2 estimation procedures.

Tournament	Division	OPR WMPR	OPR WMPRC1	OPR WMPRC2	OPRC1 WMPRC1	OPRC1 WMPRC2	OPRC2 WMPRC1	OPRC2 WMPRC2	WMPR OPRC1	WMPR OPRC2
Houston	Carver	78%	77%	78%	67%	68%	77%	78%	67%	78%
	Galileo	81%	81%	82%	81%	82%	81%	81%	82%	81%
	Hopper	77%	76%	77%	76%	77%	75%	76%	77%	76%
	Newton	79%	79%	79%	77%	77%	77%	78%	77%	77%
	Roebling	80%	80%	79%	77%	77%	78%	78%	77%	78%
	Turing	78%	78%	76%	77%	75%	78%	76%	77%	78%
Detroit	Archimedes	80%	79%	79%	78%	78%	78%	78%	79%	79%
	Carson	79%	77%	77%	71%	71%	76%	76%	73%	78%
	Curie	83%	80%	79%	80%	78%	79%	78%	83%	83%
	Daly	82%	82%	82%	77%	78%	78%	78%	77%	78%
	Darwin	79%	77%	76%	77%	76%	76%	74%	79%	77%
	Tesla	84%	81%	80%	78%	76%	77%	76%	80%	79%

robots in the Turing division, the precision between FRC top-8 robots and WMPRC1-based top-N robots is comparable to or higher than the precisions between FRC top-8 robots and WMPR-based top-N robots and between FRC top-8 robots and WMPRC2-based top-N robots. Further, the precisions between FRC top-8 robots and OPRC1-based top-N robots and between FRC top-8 robots and OPRC2-based top-N robots are higher than the precision between FRC top-8 robots and WMPRC1-based top-N robots except in the Hopper division. Evidenced by the results in Tables 13–14, the conclusion for the precision of FRC top-8 robots and model-based top-N robots can also be made for the recall of FRC top-8 robots and model-based top-N robots.

In Tables 15–16, we can observe that M_9 matches should be enough to ensure the stability of accuracies of the OPRC and WMPRC models in the qualification stage. Tables 17–18 further show that the OPRC2, WMPRC1, and WMPRC2 estimation procedures have relatively high rank correlations of estimated robot strengths on M_ℓ and $M_{\ell+1}$ matches, $\ell = 6, \ldots, 9$. Except in the Carver division, the same conclusion can be made for the OPRC1 estimation procedure. It must be emphasized that these observed results can only refer to the examined case study. According to model-based

Table 11
Precisions of FRC top-8 robots and model-based top-8 robots, which are determined by the OPR, OPRC1, OPRC2, WMPR, WMPRC1, and WMPRC2 estimation procedures.

Tournament	Division	OPR	OPRC1	OPRC2	WMPR	WMPRC1	WMPRC2
Houston	Carver	63%	100%	75%	50%	63%	63%
	Galileo	75%	75%	75%	75%	75%	75%
	Hopper	75%	75%	75%	75%	100%	75%
	Newton	88%	88%	88%	63%	75%	75%
	Roebling	63%	63%	88%	63%	63%	63%
	Turing	38%	75%	63%	50%	50%	63%
Detroit	Archimedes	75%	75%	75%	63%	63%	63%
	Carson	50%	75%	64%	50%	50%	50%
	Curie	50%	63%	63%	38%	50%	50%
	Daly	63%	88%	88%	50%	50%	50%
	Darwin	50%	63%	63%	25%	38%	25%
	Tesla	63%	88%	88%	50%	75%	50%

Table 12
Precisions of FRC top-8 robots and model-based top-16 robots, which are determined by the OPR, OPRC1, OPRC2, WMPR, WMPRC1, and WMPRC2 estimation procedures.

Tournament	Division	OPR	OPRC1	OPRC2	WMPR	WMPRC1	WMPRC2
Houston	Carver	56%	100%	56%	44%	56%	56%
	Galileo	69%	69%	69%	56%	75%	69%
	Hopper	75%	75%	75%	63%	88%	63%
	Newton	75%	81%	81%	56%	63%	63%
	Roebling	75%	81%	75%	63%	69%	69%
	Turing	69%	75%	75%	63%	69%	69%
Detroit	Archimedes	69%	75%	75%	50%	56%	56%
	Carson	75%	81%	81%	50%	50%	50%
	Curie	69%	69%	69%	50%	69%	75%
	Daly	69%	69%	69%	63%	63%	63%
	Darwin	75%	75%	75%	50%	50%	56%
	Tesla	69%	81%	94%	75%	75%	81%

Table 13
Recalls of FRC top-8 robots and model-based top-8 robots, which are determined by the OPR, OPRC1, OPRC2, WMPR, WMPRC1, and WMPRC2 estimation procedures.

Tournament	Division	OPR	OPRC1	OPRC2	WMPR	WMPRC1	WMPRC2
Houston	Carver	63%	100%	75%	50%	63%	63%
	Galileo	75%	75%	75%	75%	75%	75%
	Hopper	75%	75%	75%	75%	100%	75%
	Newton	88%	88%	88%	63%	75%	75%
	Roebling	63%	63%	88%	63%	63%	63%
	Turing	38%	75%	63%	50%	50%	63%
Detroit	Archimedes	75%	75%	75%	63%	63%	63%
	Carson	50%	75%	63%	50%	50%	50%
	Curie	50%	63%	63%	38%	50%	50%
	Daly	63%	88%	88%	50%	50%	50%
	Darwin	50%	63%	63%	25%	38%	25%
	Tesla	63%	88%	88%	50%	75%	50%

top-8 robots, which are selected according to estimated robot strengths on M matches, high rank correlations of estimated robot strengths on M_ℓ and $M_{\ell+1}$ matches, $\ell = 6, \ldots, 9$, (see Tables 19–20) are found in the Archimedes and Tesla divisions from the OPRC1 estimation procedure; in the Galileo, Turing, Archimedes, Darwin, and Tesla divisions from the OPRC2 estimation procedure; in the Carver, Newton (except for $\ell = 6$), Turing, Carson, and Daly divisions from the WMPRC1 estimation procedure; and in the Carver, Hooper, Newton, Carson, and Daly divisions from the WMPRC2 estimation procedure.

5. Conclusion and discussion

In our application to the 2018 FRC championships, the OPR and WMPR models have poor predictive performance. This is mainly because the ratio of the number of matches to the number of robots in each division is rather small and both models are highly over-parameterized. To enhance their predictive capacities, the OPRC and WMPRC models are proposed as possible avenues. In addition to generalize the model formulation of robot strengths, we do not assume any particular

Table 14
Recalls of FRC top-8 robots and model-based top-16 robots, which are determined by the OPR, OPRC1, OPRC2, WMPR, WMPRC1, and WMPRC2 estimation procedures.

Tournament	Division	OPR	OPRC1	OPRC2	WMPR	WMPRC1	WMPRC2
Houston	Carver	88%	100%	88%	63%	63%	63%
	Galileo	88%	88%	88%	75%	100%	88%
	Hopper	100%	100%	100%	75%	100%	75%
	Newton	88%	100%	100%	75%	88%	88%
	Roebling	88%	100%	88%	75%	75%	75%
	Turing	75%	88%	88%	75%	100%	88%
Detroit	Archimedes	88%	88%	88%	75%	88%	88%
	Carson	88%	100%	88%	63%	63%	63%
	Curie	88%	88%	88%	63%	88%	88%
	Daly	88%	88%	88%	75%	75%	75%
	Darwin	75%	75%	75%	50%	50%	63%
	Tesla	100%	100%	100%	88%	88%	88%

Table 15
The accuracies, which are estimated by the OPRC1 and OPRC2 estimation procedures, on $M_6, \ldots, M_{10}$ matches in the qualification stage.

Tournament	Division	OPRC1					OPRC2				
		M_6	M_7	M_8	M_9	M_{10}	M_6	M_7	M_8	M_9	M_{10}
Houston	Carver	85%	86%	86%	86%	84%	81%	85%	80%	85%	88%
	Galileo	78%	77%	84%	88%	88%	78%	82%	87%	88%	89%
	Hopper	84%	84%	86%	85%	85%	85%	84%	85%	85%	85%
	Newton	89%	84%	86%	82%	86%	89%	84%	86%	84%	86%
	Roebling	78%	75%	81%	86%	85%	78%	77%	77%	82%	86%
	Turing	88%	88%	87%	90%	90%	84%	85%	84%	88%	89%
Detroit	Archimedes	82%	84%	80%	81%	82%	87%	84%	82%	83%	82%
	Carson	71%	76%	78%	79%	81%	70%	71%	75%	79%	81%
	Curie	78%	79%	81%	80%	83%	76%	79%	82%	80%	81%
	Daly	87%	87%	86%	85%	87%	87%	87%	87%	86%	87%
	Darwin	79%	83%	82%	81%	80%	82%	83%	79%	80%	81%
	Tesla	85%	85%	88%	88%	87%	87%	85%	83%	86%	87%

Table 16
The accuracies, which are estimated by the WMPRC1 and WMPRC2 estimation procedures, on $M_6, \ldots, M_{10}$ matches in the qualification stage.

Tournament	Division	WMPRC1					WMPRC2				
		M_6	M_7	M_8	M_9	M_{10}	M_6	M_7	M_8	M_9	M_{10}
Houston	Carver	85%	86%	88%	88%	89%	87%	89%	89%	88%	88%
	Galileo	88%	84%	87%	88%	88%	85%	83%	85%	87%	88%
	Hopper	87%	90%	90%	92%	89%	82%	83%	86%	86%	89%
	Newton	95%	92%	89%	91%	92%	93%	90%	89%	92%	92%
	Roebling	89%	87%	89%	91%	90%	84%	84%	86%	91%	91%
	Turing	94%	96%	96%	94%	95%	94%	96%	96%	93%	94%
Detroit	Archimedes	84%	84%	83%	85%	88%	84%	89%	86%	85%	89%
	Carson	85%	83%	85%	86%	86%	85%	84%	84%	87%	86%
	Curie	85%	85%	88%	85%	86%	88%	84%	89%	88%	88%
	Daly	90%	92%	90%	91%	89%	95%	94%	92%	90%	91%
	Darwin	85%	86%	86%	84%	86%	84%	87%	84%	87%	85%
	Tesla	86%	91%	88%	94%	93%	91%	92%	93%	94%	93%

distribution on the underlying distributions of errors in the proposed models. In the analysis of such paired comparison data, the estimated accuracies of the WMPRC and OPRC models are about $14\% - 26\%$ and $8\% - 19\%$ higher than those of the WMPR and OPR models, respectively. It is notable that there is no need to take into account a specific-alliance advantage, a defensive component, and any dynamic effect in the WMPRC model.

As stated in the introduction, performing well in the qualification stage comes with the benefit of being able to choose your alliance for the playoff stage. Compared to the other advantage of being guaranteed a spot in the playoff stage, this benefit is often taken for granted, with some teams opting to rely on chance qualification-round synergy, published OPR rankings, or cooperation history from tournaments or seasons past. Our methodology, shown to provide better predictive capabilities with regard to the playoff rounds, outputs a more reliable ranking of robots by rating that should assist in top teams' decision making during alliance selection. As a baseball field manager might look to past data and statistical models to guide a decision on whether or not to call in a relief pitcher or substitute a pinch hitter in a crucial inning, we hope that future FRC teams may look to our model to guide their endeavors at the highest levels of play, and that such adoption leads to a greater experience and appreciation of all facets of this complicated game.

Table 17
The rank correlations of robot strengths, which are estimated by the OPRC1 and OPRC2 estimation procedures, of all robots on M_ℓ and $M_{\ell+1}$ matches, $\ell = 6, \ldots, 9$, in the qualification stage.

Tournament	Division	OPRC1				OPRC2			
		M_6&M_7	M_7&M_8	M_8&M_9	M_9&M_{10}	M_6&M_7	M_7&M_8	M_8&M_9	M_9&M_{10}
Houston	Carver	73%	73%	73%	73%	92%	92%	92%	92%
	Galileo	93%	91%	93%	93%	93%	92%	93%	93%
	Hopper	90%	90%	90%	90%	89%	89%	89%	89%
	Newton	89%	89%	89%	89%	92%	92%	92%	92%
	Roebling	90%	90%	90%	90%	92%	92%	92%	92%
	Turing	91%	92%	92%	92%	93%	93%	95%	95%
Detroit	Archimedes	94%	94%	94%	94%	93%	93%	93%	93%
	Carson	83%	83%	83%	83%	87%	93%	94%	94%
	Curie	94%	94%	95%	95%	94%	94%	95%	96%
	Daly	85%	85%	85%	85%	86%	86%	86%	86%
	Darwin	93%	93%	93%	93%	92%	92%	92%	92%
	Tesla	87%	87%	87%	87%	85%	85%	85%	85%

Table 18
The rank correlations of robot strengths, which are estimated by the WMPRC1 and WMPRC2 estimation procedures, of all robots on M_ℓ and $M_{\ell+1}$ matches, $\ell = 6, \ldots, 9$, in the qualification stage.

Tournament	Division	WMPRC1				WMPRC2			
		M_6&M_7	M_7&M_8	M_8&M_9	M_9&M_{10}	M_6&M_7	M_7&M_8	M_8&M_9	M_9&M_{10}
Houston	Carver	92%	92%	92%	92%	95%	95%	95%	95%
	Galileo	90%	90%	90%	90%	92%	92%	92%	92%
	Hopper	91%	91%	91%	91%	96%	95%	97%	97%
	Newton	96%	96%	96%	96%	96%	96%	96%	96%
	Roebling	93%	93%	93%	93%	94%	94%	94%	94%
	Turing	92%	92%	93%	93%	90%	90%	90%	90%
Detroit	Archimedes	94%	94%	95%	95%	88%	88%	88%	88%
	Carson	90%	90%	90%	90%	94%	94%	94%	94%
	Curie	89%	89%	89%	89%	87%	87%	87%	87%
	Daly	94%	94%	94%	94%	94%	96%	96%	96%
	Darwin	92%	92%	92%	92%	91%	91%	91%	91%
	Tesla	89%	89%	89%	89%	88%	88%	88%	88%

Table 19
The rank correlations of robot strengths, which are estimated by the OPRC1 and OPRC2 estimation procedures, of model-based top-8 robots on M_ℓ and $M_{\ell+1}$ matches, $\ell = 6, \ldots, 9$, in the qualification stage.

Tournament	Division	OPRC1				OPRC2			
		M_6&M_7	M_7&M_8	M_8&M_9	M_9&M_{10}	M_6&M_7	M_7&M_8	M_8&M_9	M_9&M_{10}
Houston	Carver	77%	77%	77%	77%	80%	80%	80%	80%
	Galileo	89%	89%	79%	89%	84%	84%	84%	84%
	Hopper	63%	63%	63%	63%	63%	63%	63%	63%
	Newton	63%	63%	63%	63%	63%	63%	63%	63%
	Roebling	71%	71%	71%	71%	71%	71%	71%	71%
	Turing	71%	71%	71%	71%	88%	88%	88%	88%
Detroit	Archimedes	84%	84%	84%	84%	88%	88%	88%	88%
	Carson	50%	50%	50%	50%	80%	80%	80%	80%
	Curie	86%	75%	71%	89%	86%	71%	75%	89%
	Daly	79%	79%	79%	79%	79%	79%	79%	79%
	Darwin	70%	73%	73%	73%	84%	84%	84%	84%
	Tesla	91%	91%	91%	91%	84%	84%	84%	84%

To sum up, our major contributions include:

1. Existing models are extended to more general models, which have better predictive performance in our application, with latent clusters of robot strengths;
2. Effective estimation procedures, in which the estimation problem is transferred to the model selection problem, are developed to simultaneously estimate the number of clusters, clusters of robots, and robot strengths;
3. A very flexible semiparametric regression model is proposed to predict a future match outcome; and
4. The stability of estimated robot strengths and accuracies is investigated to determine an appropriate number of matches in the qualification stage.

Table 20
The rank correlations of robot strengths, which are estimated by the WMPRC1 and WMPRC2 estimation procedures, of model-based top-8 robots on M_ℓ and $M_{\ell+1}$ matches, $\ell = 6, \ldots, 9$, in the qualification stage.

Tournament	Division	WMPRC1				WMPRC2			
		M_6&M_7	M_7&M_8	M_8&M_9	M_9&M_{10}	M_6&M_7	M_7&M_8	M_8&M_9	M_9&M_{10}
Houston	Carver	88%	88%	88%	88%	91%	91%	91%	91%
	Galileo	71%	71%	71%	71%	71%	71%	71%	71%
	Hopper	80%	80%	80%	80%	96%	89%	96%	96%
	Newton	80%	88%	88%	88%	88%	88%	88%	88%
	Roebling	80%	80%	80%	80%	80%	80%	80%	80%
	Turing	88%	88%	88%	88%	79%	79%	79%	79%
Detroit	Archimedes	68%	75%	82%	93%	80%	80%	80%	80%
	Carson	84%	84%	84%	84%	84%	84%	84%	84%
	Curie	63%	63%	63%	63%	73%	73%	73%	73%
	Daly	86%	89%	86%	89%	89%	89%	89%	93%
	Darwin	73%	73%	73%	73%	73%	73%	73%	73%
	Tesla	77%	77%	77%	77%	63%	63%	63%	63%

Some measures are further used to assess the predictive ability of competing models and the agreement related to FRC ratings and model-based robot strengths. With slight modifications, our methodology should be successfully applied to other team games.

By taking into account random robot strengths, i.e. $\beta_i = \beta^{c_0}_{g_i} + \varepsilon_{g_i i}$ with $\varepsilon_{g_i i}$ having mean zero and variance $\sigma^2_{g_i}$, $i = 1, \ldots, K$, the proposed fixed effects model in (12) can be further generalized to the following random effects model:

$$Y = X^{c_0}\beta^{c_0} + \left(\sum_{c=1}^{c_0}\sum_{j=1}^{K_c}\varepsilon_{cj}X_{(cj)} + \varepsilon\right), \tag{26}$$

where $X_{(cj)}$ is the designed covariate vector of $X_{(i)}$ with $g_i = c$, $i = 1, \ldots, K$, ε_{cj}, which has mean zero and variance σ^2_c, $c = 1, \ldots, c_0$, $j = 1, \ldots, K_c$, are mutually independent, and ε is independent of ε_{cj}'s. Obviously, the WMPRR model is a special case of the above model. Different from the error in model (12), the error term $\left(\sum_{c=1}^{c_0}\sum_{j=1}^{K_c}\varepsilon_{cj}X_{(cj)} + \varepsilon\right)$ in model (26) is correlated. In particular, the model formulation avoids the problem of ties in β_i's in model (12). It will be challenging to develop an appropriate estimation procedure for such a random effects model, especially for the determination of the number of clusters and clusters of robots. An investigation for its predictive ability would be worthwhile in future research. By making distributional assumptions on ε_{cj}'s and ε, existing methods in the literature, such as Laird and Ware (1982) and Diggle et al. (2002), can also be used to estimate match-specific effects as well as robot-specific effects on each match score.

We note that no individual robot characteristic (e.g. the speed, weight, or height of a robot), which is expected to bring a contribution to alliance scores, is considered in existing models. Another factor that affects match scores is penalty scores gained from opposing robots who violate game rules. For some FRC games, obstacles in the middle of the field change match-by-match relying on the audience selection. Since some robots might have an advantage on the designed obstacles in their matches, such an environmental factor needs to be carefully formulated. The winning rate of each robot in its former matches should be also helpful in the model development. As emphasized in this article, robots in the playoff stage are not randomly assigned to alliances and matches. A consideration of the above confounders is expected to enhance the predictive power of the current models. With this in mind, our goal is to continue improving the proposed model for the convenience of future robotics teams, leading to high quality competitions.

Acknowledgments

Chin-Tsang Chiang's research was partially supported by the Ministry of Science and Technology grant 109-2118-M-002-002-MY2 (Taiwan). Alejandro Lim completed a significant portion of the research work during his senior year at the Westminster Schools (Atlanta) and while visiting the Institute of Applied Mathematical Sciences, National Taiwan University as an intern. The authors are grateful to Dr. Alvin Lim for useful discussions. The authors would also like to thank the editor, the associate editor, and two reviewers for their constructive comments.

Appendix

Let $H^c = X^c\left(X^{c\top}X^c\right)^{-1}X^{c\top}$ and $\bar{H}^c = \bar{X}^c\left(\bar{X}^{c\top}\bar{X}^c\right)^{-1}\bar{X}^{c\top}$. For the OPRC model, we derive that

$$\widehat{\beta}^c_{-s} = \widehat{\beta}^c - \frac{\left(X^{c\top}X^c\right)^{-1}\left(\left(1 - H^c_{M+sM+s}\right)e^c_s + H^c_{M+ss}e^c_{M+s}\right)X_s}{\left(1 - H^c_{ss}\right)\left(1 - H^c_{M+sM+s}\right) - H^{c2}_{M+ss}}$$

$$-\frac{\left(X^{c\top}X^{c}\right)^{-1}\left(\left(1-H^{c}_{ss}\right)e^{c}_{M+s}+H^{c}_{M+ss}e^{c}_{s}\right)X_{M+s}}{\left(1-H^{c}_{ss}\right)\left(1-H^{c}_{M+sM+s}\right)-H^{c2}_{M+ss}},\ s=1,\ldots,M. \tag{A.1}$$

In light of the above result, $\widehat{F}^{c}_{-s}(v)$ has the form

$$\widehat{F}^{c}_{-s}(v)=\frac{1}{M-1}\sum_{s_1\neq s}I\left(\left(Y_{M+s_1}-Y_{s_1}\right)-\left(X^{c}_{M+s_1}-X^{c}_{s_1}\right)^{\top}\widehat{\beta}^{c}_{-s}\leq v\right) \tag{A.2}$$

with

$$\begin{aligned}\left(X^{c}_{M+s_1}-X^{c}_{s_1}\right)^{\top}\widehat{\beta}^{c}_{-s}=\left(X^{c}_{M+s_1}-X^{c}_{s_1}\right)^{\top}\widehat{\beta}^{c}&-\frac{\left(H^{c}_{sM+s_1}-H^{c}_{ss_1}\right)\left(\left(1-H^{c}_{M+sM+s}\right)e^{c}_{s}+H^{c}_{M+ss}e^{c}_{M+s}\right)}{\left(1-H^{c}_{ss}\right)\left(1-H^{c}_{M+sM+s}\right)-H^{c2}_{M+ss}}\\&-\frac{\left(H^{c}_{M+sM+s_1}-H^{c}_{M+ss_1}\right)\left(\left(1-H^{c}_{ss}\right)e^{c}_{M+s}+H^{c}_{M+ss}e^{c}_{s}\right)}{\left(1-H^{c}_{ss}\right)\left(1-H^{c}_{M+sM+s}\right)-H^{c2}_{M+ss}},\ s,s_1=1,\ldots,M.\end{aligned} \tag{A.3}$$

As for the WMPRC model, the properties in equation (8.5) of Sen and Srivastava (1990) and Remark 4 enable us to have

$$\widehat{\beta}^{*c}_{-s}=\widehat{\beta}^{*c}-\frac{\left(\bar{X}^{c\top}\bar{X}^{c}\right)^{-1}\bar{X}^{c}_{s}e^{c}_{s}}{\left(1-\bar{H}^{c}_{ss}\right)}\ \text{and}\ X^{c\top}_{s_1}\widehat{\beta}^{c}_{-s}=\bar{X}^{c\top}_{s_1}\widehat{\beta}^{*c}_{-s},\ s,s_1=1,\ldots,M. \tag{A.4}$$

It follows that

$$\widehat{F}^{c}_{-s}(v)=\frac{1}{M-1}\sum_{s_1\neq s}I\left(Y_{s_1}-X^{c\top}_{s_1}\widehat{\beta}^{c}_{-s}\leq v\right)\ \text{with}\ X^{c\top}_{s_1}\widehat{\beta}^{c}_{-s}=\bar{X}^{c\top}_{s_1}\widehat{\beta}^{*c}-\frac{\bar{H}^{c}_{s_1s}e^{c}_{s}}{\left(1-\bar{H}^{c}_{ss}\right)},\ s,s_1=1,\ldots,M. \tag{A.5}$$

References

Akaike, H., 1974. A new look at statistical model identification. IEEE Trans. Automat. Control 19 (6), 716–723.
Aldous, D., 2017. Elo ratings and the sports model: a neglected topic in applied probability?. Stat. Sci. 32 (4), 616–629.
Bradley, R.A., 1953. Some statistical methods in taste testing and quality evaluation. Biometrics 9 (1), 22–38.
Bradley, R.A., Terry, M.E., 1952. Rank analysis of incomplete block designs: I. the method of paired comparisons. Biometrika 39 (3/4), 324–345.
Cattelan, M., Varin, C., Firth, D., 2013. Dynamic Bradley-Terry modelling of sports tournaments. J. R. Stat. Soc. Ser. C 62 (1), 135–150.
Chen, Z., Sun, Y., El-nasr, M.S., Nguyen, T.D., 2017. Player skill decomposition in multiplayer online battle arenas. arXiv:1702.06253.
Chiang, C.-T., Chiu, C.-H., 2012. Nonparametric and semiparametric optimal transformations of markers. J. Multivariate Anal. 103 (1), 124–141.
Clark, N., Macdonald, B., Kloo, I., 2020. A Bayesian adjusted plus-minus analysis for the esport Dota 2. http://dx.doi.org/10.1515/jqas-2019-0103.
Collins, L.M., Lanza, S.T., 2010. Latent Class and Latent Transition Analysis: with Applications in the Social, Behavioral, and Health Sciences. Wiley, New York.
Deshpande, S.K., Jensen, S.T., 2016. Estimating an NBA players impact on his teams chances of winning. J. Quant. Anal. Sports 12 (2), 51–72.
Diggle, P.J., Heagerty, P., Liang, K.Y., Zeger, S.L., 2002. Analysis of Longitudinal Data. Oxford, Oxford.
Elo, A.E., 1978. The Rating of Chess Players, Past and Present. Arco, New York.
Fahrmeir, L., Tutz, G., 2001. Multivariate Statistical Modelling Based on Generalized Linear Models. Springer, New York.
Fang, E., 2017. The math behind OPR – an introduction. The Blue Alliance Blog. https://blog.thebluealliance.com/2017/10/05/the-math-behind-opr-an-introduction/.
Feamhead, P., Taylor, B., 2011. On estimating the ability of NBA players. J. Quant. Anal. Sports 7 (3).
FIRST, 2017a. 2017 FIRST robotics competition STEAMWORKS game animation. https://www.youtube.com/watch?v=EMiNmJW7enI.
FIRST, 2017b. 2017 game and season manual. https://firstfrc.blob.core.windows.net/frc2017/Manual/2017FRCGameSeasonManual.pdf.
Gardner, W., 2015. An overview and analysis of statistics used to rate FIRST robotics teams. https://www.chiefdelphi.com/uploads/default/original/3X/e/b/ebd6f208f128bcc2d44267c7b6d25a4ae53aa8a4.pdf.
Giraud, C., 2015. Introduction to High-Dimensional Statistics. Chapman & Hall/CRC, Florida.
Goodman, L.A., 1974. Exploratory latent structure analysis using both identifiable and unidentifiable models. Biometrika 61, 215–231.
Gordon, A.D., 1987. A review of hierarchical classification. J. R. Stat. Soc. A 150 (2), 119–137.
Han, A.K., 1987. Non-parametric analysis of a generalized regression model: the maximum rank correlation estimator. J. Econom. 35 (2), 303–316.
Hass, Z., Craig, B., 2018. Exploring the potential of the plus/minus in NCAA women's volleyball via the recovery of court presence information. J. Sports Anal. 4, 285–295.
Hochberg, Y., Tamhane, A.C., 1987. Multiple Comparison Procedures. Wiley, New York.
Hothorn, T., Bretz, F., Westfall, P., 2008. Simultaneous inference in general parametric models. Biom. J. 50 (3), 346–363.
Hsu, J.C., 1996. Multiple comparisons: theory and methods. Chapman & Hall, London.
Huang, T.K., Lin, C.J., Weng, R.C., 2008. Ranking individuals by group comparisons. J. Mach. Learn. Res. 9, 2187–2216.
Hvattum, L.M., 2019. A comprehensive review of plus-minus ratings for evaluating individual players in team sports. Int. J. Comput. Sci. Sports 18 (1), 1–23.
Ilardi, S., Barzilai, A., 2008. Adjusted plus-minus ratings: new and improved for 2007-2008. https://www.82games.com/ilardi2.htm.
Islam, M.M., Khan, J.R., Raheem, E., 2017. Bradley-Terry model for assessing the performance of ten ODI cricket teams adjusting for home ground effect. Data Sci. J. 16, 657–668.
Kendall, M., 1938. A new measure of rank correlation. Biometrika 30 (1/2), 81–89.
Kovalchik, S., 2020. Extension of the Elo rating system to margin of victory. Int. J. Forecast. 36, 1329–1341.
Krzanowski, W.J., Lai, Y.T., 1985. A criterion for determining the number of clusters in a data set. Biometrics 44, 23–34.
Laird, N.M., Ware, J.H., 1982. Random-effects models for longitudinal data. Biometrics 38 (4), 963–974.
Law, E., 2008. New scouting database from team 2834. https://www.chiefdelphi.com/forums/showpost.php?p=835222&postcount=48.
Lazarsfeld, P.F., Henry, N.W., 1968. Latent Structure Analysis. Houghton Mifflin, Boston.
Macdonald, B., 2011. A regression-based adjusted plus-minus statistic for NHL players. J. Quant. Anal. Sports 7 (3).

Mammen, E., 1996. Empirical process of residuals for high-dimensional linear models. Ann. Statist. 24 (1), 307–325.
Marx, I., 1962. Transformation of series by a variant of Stirling's numbers. Am. Math. Mon. 69 (6).
Mosteller, F., 1951. Remarks on the method of paired comparisons: I. the least squares solution assuming equal standard deviations and equal correlations. Psychometrika 16 (1), 3–9.
Perry, J.W., Kent, A., Berry, M.M., 1955. Machine literature searching X. Machine language; factors underlying its design and development. Am. Doc. 6 (4), 242–254.
Portnoy, S., 1984. Asymptotic behavior of M-estimators of p regression parameters when p^2/n is large. I. Consistency. Ann. Statist. 12, 1298–1309.
Rosenbaum, D.T., 2004. Measuring how NBA players help their teams win. https://www.82games.com/comm30.htm.
Sæbø, O.D., Hvattum, L.M., 2015. Evaluating the efficiency of the association football transfer market using regression based player ratings. In: NIK-2015 Conference, Vol. 4. Bibsys Open Journal Sysytems, p. 12.
Salmeri, A., 1962. Introduzione alla teoria dei coefficienti fattoriali. G. Mat. Battagl. 90 (10), 44–54.
Schuckers, M.E., Lock, D.F., Wells, C., Knickerbocker, C.J., Lock, R.H., 2011. National Hockey League skater ratings based upon on-ice events: an adjusted minus/plus probability (AMPP) approach. https://www.researchgate.net/profile/Michael_Schuckers/publication/266493342_Copyright_C_2011_Michael_Schuckers_National_Hockey_League_Skater_Ratings_Based_upon_All_On-Ice_Events_An_Adjusted_MinusPlus_Probability_AMPP_Approach/links/55251f210cf2caf11bfd146c.pdf.
Schwarz, G., 1978. Estimating the dimension of a model. Ann. Statist. 6 (2), 461–464.
Sen, A., Srivastava, M., 1990. Regression analysis: theory, methods, and applications. Springer, Berlin.
Sill, J., 2010. Improved NBA adjusted +/- using regularization and out-of-sample testing. In: MIT Sloan Sports Analytics Conference. MIT.
Sismanis, Y., 2010. How i won the "Chess Ratings - Elo vs the Rest of the World" competition. arXiv:1012.4571.
Slowinski, S., 2010. What is WAR?. https://library.fangraphs.com/misc/war/.
Sokal, R.R., Michener, C.D., 1958. A Statistical Method for Evaluating Systematic Relationships. (38), pp. 1409–1438,
Spearman, C., 1904. The proof and measurement of association between two things. Am. J. Psychol. 15 (1), 72–101.
Sugar, C., James, G., 2003. Finding the number of clusters in a data set: an information theoretic approach. J. Amer. Statist. Assoc. 98 (463), 750–763.
TeamRankings, 2019. NBA team possessions per game. https://www.teamrankings.com/nba/stat/possessions-per-game.
The Blue Alliance, 2018. Hopper division 2018. https://www.thebluealliance.com/event/2018hop.
Thurstone, L.L., 1927. A law of comparative judgment. Psychol. Rev. 34 (4), 273–286.
Tibshirani, R., Walther, G., Hastie, T., 2001. Estimating the number of clusters in a data set via the gap statistic. J. R. Stat. Soc. Ser. B Stat. Methodol. 63 (2), 411–423.
Vecchio, T.J., 1966. Predictive value of a single diagnostic test in unselected populations. New Engl. J. Med. 274 (21), 1171–1173.
Wang, J., 2010. Consistent selection of the number of clusters via crossvalidation. Biometrika 97 (4), 893–904.
Weingart, S., 2006. Offense/defense rankings for 1043 teams. https://www.chiefdelphi.com/forums/showpost.php?p=484220&postcount=19.
Weng, R.C., Lin, C.J., 2011. A Bayesian approximation method for online ranking. J. Mach. Learn. Res. 12, 267–300.
Wikipedia, 2018. FIRST robotics competition. Retrieved August 30, 2018 from. https://en.wikipedia.org/wiki/FIRST_Robotics_Competition.
Woolner, K., 2001a. Introduction to VORP: value over replacement player. https://web.archive.org/web/20070928064958/http://www.stathead.com/bbeng/woolner/vorpdescnew.htm.
Woolner, K., 2001b. VORP: measuring the value of a baseball player's performance. https://web.archive.org/web/20070928064958/http://www.stathead.com/bbeng/woolner/vorpdescnew.htm.
Yerushalmy, J., 1947. Statistical problems in assessing methods of medical diagnosis with special reference to x-ray techniques. Public Health Rep. 62 (2), 1432–1439.