

Net best-ball team composition in golf

Yifan Wu^a, Peter Chow-White^b and Tim B. Swartz^{a,1,*}

^a*Department of Statistics and Actuarial Science, Burnaby BC, Canada*

^b*School of Communication, Simon Fraser University, Burnaby BC, Canada*

Abstract. This paper proposes simple methods of forming two-player and four-player golf teams for the purposes of net best-ball tournaments in stroke play format. The proposals are based on the recognition that variability is an important consideration in team composition; highly variable players contribute greatly in a best-ball setting. A theoretical rationale is provided for the proposed team formations. In addition, simulation studies are carried out which compare the proposals against other common methods of team formation. In these studies, the proposed team compositions lead to competitions that are more fair.

Keywords: Distribution of minimum, Handicapping in golf, Order statistics, Simulation

1. Introduction

One of the compelling features of golf is that players of vastly different abilities can compete against one another and have a “fair” match. This is accomplished by the handicapping system which has a long history of refinements extending back to the 1600’s (Yun 2011a, 2011b, 2011c, 2011d).

There is a considerable literature on golf handicapping, and the consensus is that most handicapping systems provide a modest advantage to the stronger player in both stroke and match play formats involving two players (Chan, Madras and Puterman 2018, Kupper et al. 2012, Bingham and Swartz 2000, Scheid 1977 and Pollock 1974). In fact, Section 10-2 of the United States Golf Association (USGA) Handicap System Manual (USGA 2016) states that the handicap formula provides an “incentive for players to improve their golf games” whereby a small bonus for excellence advantage is given to the stronger player.

However, in competitions involving multiple players, varying rules and team formats, fairness may be greatly violated. For example, Bingham and Swartz (2000) suggested that weaker golfers have a considerable advantage winning a tournaments based on net scores. Grasman and Thomas (2013) investigated *scramble* competitions and provided suggestions for assigning teams. And of particular relevance to this paper, Hurley and Sauerbrei (2015) demonstrated that team net best-ball matches are not generally fair.

Handicapping in golf takes different forms depending on the governing body. In this paper, we focus on the handicapping system used by the USGA and the Royal Canadian Golf Association (RCGA). And although golf is a stochastic game, the USGA/RCGA handicapping system was not developed using the tools of probability theory. On the other hand, we make use of the stochastic nature of golf to provide team compositions in net best-ball competitions that are more fair than the status quo. For the purposes of this paper, we define a fair system involving n teams as one where the probability of each team finishing in j th place is $1/n$, $j = 1, \dots, n$. Surprisingly, although fairness in golf is a much discussed topic, the above definition does not appear to exist in the golf literature.

There are two related papers that concern the problem of team formation in net best-ball compe-

¹Swartz has been partially supported by the Natural Sciences and Engineering Research Council of Canada (NSERC). The authors thank two anonymous reviewers whose comments have helped strengthen the paper.

*Corresponding author: Tim B. Swartz, Department of Statistics and Actuarial Science, Simon Fraser University, 8888 University Drive, Burnaby BC, Canada V5A1S6. E-mail: tim@stat.sfu.ca.

titions. Siegbahn and Hearn (2010) studied *fourball*; a two-player versus two-player event where handicapping is used. Like Bingham and Swartz (2000), golfer variability was a prominent focus of their study where the variability of golfer performance was parametrized and estimated as a function of handicap. Siegbahn and Hearn (2010) concluded that high handicap golfers (i.e., weak golfers) have an advantage in fourball, and they suggested tie-breaking rules to reduce the unfairness. As discussed in Siegbahn and Hearn (2010), previous studies on fairness in fourball matches focused on the difference in handicaps between teammates as a predictor of fourball success.

Pavlikov, Hearn and Uryasev (2014) built on the results of Siegbahn and Hearn (2010) to specifically address team composition in net best-ball tournament settings. They developed a sophisticated search algorithm over the combinatorial space of potential team compositions. Optimal team formations were sought in the sense that all teams have nearly the same probability of winning. When the number of golfers $n < 40$, it was asserted that the program can be run in reasonable computational times. A feature of the approach proposed by Pavlikov, Hearn and Uryasev (2014) is that the algorithm is applicable to any prescribed team size. A drawback of the approach involves the reliance on tables that provide average scoring distributions for players of a given handicap. An implication of the use of the tables is an imposed monotonicity between handicap and performance variability. That is, the table exhibits increasing scoring variability with increasing handicap. Whereas it is generally the case that high handicap golfers tend to be more variable, there are clearly instances of high handicap golfers who are consistent. For example, imagine a senior golfer who does not hit the ball far, is straight off the tee and rarely gets into trouble (i.e., does not land in the rough, hazards, water, etc.). It is the third author's experience that such golfers do exist. Moreover, the dataset which we consider in Section 4.2 suggests there is not a strictly monotonic relationship between handicap and variability. Following Swartz (2009), we estimate variability individually for golfers, and this forms the critical component for our team formation proposals in net best-ball tournaments. And importantly, the proposed estimation of golfer specific variability is a straightforward side calculation of handicap.

In Section 2, we review various background material that is related to the development of team composition. This includes details concerning the

rules related to net best-ball tournaments, the current handicapping system and related literature. In Section 3, our proposals for team composition are developed. They are based on the recognition that variability is an important consideration in terms of player performance in net best-ball competitions. The basic idea is that golfers of high variability are matched up with golfers of low variability. Matching procedure are developed for both two-man and four-man team competitions. A noteworthy aspect of the procedures is that they do not require sophisticated software and are simple to implement. A theoretical justification is given for the proposed methods of team formation. In Section 4, two simulation studies are provided. The first study is based on a theoretical model for golf scores and investigates the performance of all possible team compositions. The second study is based on a resampling procedure of actual golf scores and investigates common practices involving team composition. In both studies, it is demonstrated that the proposed methods of team composition lead to competitions that are more fair. We conclude with a short discussion in Section 6.

Although the paper is written in a theoretical style, the results can be condensed into some straightforward non-technical advice that has wide applicability. For example, in golf, it is common for foursomes of friends to meet on the first tee and decide to play net best-ball matches, two players versus two players. In this case, how should the teams be formed? The simple answer is that matches are fairest if the most consistent player is paired with the least consistent player. Friends often know who is consistent and who is not.

There is also a strategic aspect to our results. For example, suppose that in an important competition (e.g. the Ryder Cup or the Presidents Cup) there is a *four ball* component to the event. Of course, in such competitions, there is no handicapping involved. However, if the players on a team are of nearly equal ability, then issues of consistency may be taken into account. In some circumstances, it may be thoughtful to pair a long hitter capable of many birdies (e.g. Bubba Watson) with a shorter hitter who is known for "grinding" (e.g. Jim Furyk). Alternatively, a captain may want to pair two players who are inconsistent but are "birdie machines" to form a formidable pairing.

2. Background material

Although the proposed method for net best-ball team composition is easy to describe and to

implement, some background material needs to be introduced. The background material provides the theoretical structure for the proposal.

2.1. Net best-ball competitions

Net best-ball competitions are typically based on teams of size $m = 2$ or teams of size $m = 4$. And in such competitions, we denote that there are $n \geq 2$ teams.

On the j th hole of the course, $j = 1, \dots, 18$, the i th player on a given team has a *gross score* X_{ij} which represents the number of strokes that it took to hole out. Associated with the i th player on the j th hole is a handicap allowance $h_{ij} = 0, 1, 2$ which is related to the quality of the player. The larger the value of h_{ij} , the weaker the player. Under this framework, the i th player has the resultant *net score*

$$Y_{ij} = X_{ij} - h_{ij} \quad (1)$$

on hole j . The player's team then records their *net best-ball score* on the j th hole

$$T_j = \min_{i=1, \dots, m} Y_{ij} \quad (2)$$

and the team's overall performance is based on their aggregated net best-ball score

$$T = \sum_{j=1}^{18} T_j. \quad (3)$$

The teams in the competition are then ranked according to (3) where the winning team has the lowest value of T . Various procedures exist for breaking ties. For example, with multiple ties, the team having done best on the 18th hole may be determined the winner. If a tie still exists, the criteria may then be applied to the 17th hole, then the 16th hole, etc, until the tie is broken.

The above format is known as *stroke play* which is the focus of our investigation. When there are only $n = 2$ teams, then *match play* competitions are possible. In match play, T_j is calculated as in (2), and the team with the lower value of T_j is said to have won the j th hole. The team with the greatest number of winning holes is the match play winner. Match play typically involves a competition between two teams. In this paper, we focus on the stroke play format.

2.2. The current handicapping system

Section 10 of the USGA Handicap System Manual (USGA 2016) provides the intricate details involving the calculation of handicap. However, for ease of exposition, we provide a description of the standard calculation which applies to most golfers.

Consider then a golfer's most recent 20 rounds of golf where each round is completed on a full 18-hole golf course. The k th round yields the *differential* D_k which is obtained by

$$D_k = (\text{adjusted gross score} - \text{course rating}) \\ *113/(\text{slope rating}). \quad (4)$$

In (4), the *adjusted gross score* is the player's actual score reduced according to *equitable stroke control* (ESC) which is a mechanism for limiting high scores on individual holes. The intuition is that handicap reflects potential and should not be distorted by unusually poor results. The *course rating* describes the difficulty of the course from the perspective of a *scratch* (expert) golfer. Typically, course ratings are close to the *par* score of the course where values less than (greater than) *par* indicates less (more) difficult courses. Course ratings are reported to one decimal place. The *slope rating* describes the difficulty of the course from the perspective of non-scratch golfers where a slope rating less than (greater than) 113 indicates an easier (more difficult) course than average. Slope ratings are integer-valued and lie in the interval (55, 155). The main takeaway from (4) is that large differentials correspond to poor rounds of golf and small differentials correspond to good rounds of golf. It is even possible for differentials to be negative which correspond to excellent rounds of golf. Differentials are rounded to the first decimal place.

Given a golfer's scoring record, the golfer's *handicap index* is calculated by taking 96% of the average of the 10 best (lowest) differentials and truncating the result to the first decimal place. In Section 2.3, we will see that it is instructive to write the handicap index as

$$I = (0.96)(D_{(1)} + D_{(2)} + \dots + D_{(10)})/10 \\ = (0.096)D_{(1)} + (0.096)D_{(2)} + \dots + (0.096)D_{(10)} \quad (5)$$

where $D_{(i)}$ denotes the i th order statistic of the differentials. The handicap index is the summary statistic that is used in USGA handicapping; strong golfers have small handicap indices whereas weak golfers have large handicap indices. It is possible that a golfer holds a handicap index $I < 0$, and these golfers

(mostly professionals) are referred to as *plus* golfers. The maximum allowable handicap index for men is 36.4. For many golfers, the calculation of the handicap index I is viewed as a black-box procedure. Under the RCGA jurisdiction (Golf Canada 2016), handicap index is referred to as *handicap factor*.

Recognizing that courses are of varying difficulty, the last step for the implementation of handicap involves converting the handicap index to strokes for a particular course. For a course with slope rating S , the *course handicap* for a golfer with handicap index I is given by $C = I \times S / 113$ rounded to the nearest integer.

In the context of net best-ball competitions, the course handicaps C of the players in the tournament are then used to determine the hole-by-hole handicap allowances h_{ij} in (1). It is at this point where there is some variation in how h_{ij} is obtained. According to Section 9-4(bii) of the USGA Handicap System Manual (USGA 2016), the recommended way is to first reduce the individual course handicaps C by a factor of 90%, rounding to the nearest integer. An adjustment is then made to the reduced course handicaps where the course handicap for a given golfer is set to the offset between their course handicap and the lowest (best) course handicap in the competition. For example, suppose that the best golfer in the competition has a reduced course handicap $C_{i_1} = 3$ and that some other golfer has a reduced course handicap $C_{i_2} = 24$. Then the two course handicaps are converted to $C_{i_1} = 3 - 3 = 0$ and $C_{i_2} = 24 - 3 = 21$, respectively. Then, we note that the holes on a golf course are assigned a *hole handicap* according to a stroke allocation table. The table consists of a permutation of the integers 1 to 18 where it is typically thought that increasing numbers correspond to decreasing difficulty of the holes. Denote the hole handicap on the j th hole by HDCP_j . Under the complicated framework described above, h_{ij} is determined as follows:

$$h_{ij} = \begin{cases} 0 & \text{HDCP}_j > C_i, & C_i \leq 18 \\ 1 & \text{HDCP}_j \leq C_i, & C_i \leq 18 \\ 1 & \text{HDCP}_j > C_i - 18, & C_i > 18 \\ 2 & \text{HDCP}_j \leq C_i - 18, & C_i > 18 \end{cases} \quad (6)$$

Although (6) may be difficult to digest, the idea is that relative to the strongest player, an individual with $C \leq 18$ receives a single stroke on the most difficult holes up to his handicap offset. If his handicap offset exceeds 18, then he receives two strokes on the more difficult holes and one stroke on the remaining holes.

For example, if $C = 21$, the weaker player receives two shots on handicap holes #1, #2 and #3, and one shot on the remaining 15 holes.

Chan, Madras and Puterman (2018) investigated how alternative permutations of HDCP and other innovations affect the fairness of net match play events between two players.

2.3. Related literature and ideas

In consultation with the RCGA, Swartz (2009) proposed an alternative handicapping system with the following features:

- the system retains the well-established concepts of course rating and slope rating
- the system provides a modified handicap index/factor referred to as the *mean* which has a clear interpretation in terms of actual golf performance; this is contrasted with the index/factor whose interpretation is allegedly related to potential
- the system was developed using probability theory, leading to net competitions that are more fair

The key component of the system developed by Swartz (2009) was that it incorporated variability in handicapping. And in the context of net best-ball tournaments, it is clear that amongst two golfers with the same handicap index, a highly variable golfer is more valuable to a team than a consistent golfer. For example, the highly variable golfer will obtain more net birdies which contribute positively to the overall net score of his team. On the other hand, when this highly variable golfer scores net double bogeys, these poor scores are not likely to penalize his team in the best-ball format.

As an alternative to the handicap index/factor, Swartz (2009) defined two statistics that characterize player performance. These statistics are referred to as the mean $\hat{\mu}$ and the spread $\hat{\sigma}$, and their calculation is analogous to (5). Specifically,

$$\hat{\mu} = w_1 D_{(1)} + w_2 D_{(2)} + \dots + w_{16} D_{(16)} \quad (7)$$

and

$$\hat{\sigma} = q_1 D_{(1)} + q_2 D_{(2)} + \dots + q_{16} D_{(16)} \quad (8)$$

where the weights w_i in (7) and q_i in (8) provide best linear unbiased estimators (BLUEs) of the mean and the standard deviations of the differentials where the differentials are assumed to be realizations of

independent and identically distributed normal random variables. Whereas (5) is based on 10 order statistics, (7) and (8) are based on 16 order statistics; the rationale was that data is informative and it is wasteful to discard observations. On the other hand, there is evidence that the largest differentials may not arise from a normal distribution as the true underlying distribution may be positively skewed (Siegbahn and Hearn 2010).

For the purposes of this paper, the spread $\hat{\sigma}$ in (8) plays a primary role and we record the weights q_i in Table 1. Alternative weights are recorded in Swartz (2009) when a golfer has played fewer than 20 complete rounds. When the spread calculation $\hat{\sigma}$ falls outside of the interval (1.5,8.0), it is set equal to the corresponding endpoint.

A point that is worth emphasizing is that the calculation of $\hat{\sigma}$ in (8) is simple and is directly analogous to the calculation of (5) which is part of the current handicapping system. In Section 3, we assume that the values $\hat{\sigma}$ are available for each golfer in the net best-ball competition. Moreover, the values $\hat{\sigma}$ are the only values that are needed to form teams according to our proposal. Thus, teams are formed based on the basis of individual performance variability $\hat{\sigma}$, and this is the main message of Section 2.

3. Team composition

Suppose that the number of golfers in a net best-ball tournament is an even number. The task with teams of size $m = 2$ is to pair players in a fair manner. Following (1), we let Y_{ij} denote the net score of golfer i on hole j . Although golf scores are discrete, we assume

$$Y_{ij} \sim \text{Normal}(\mu_{ij}, \tau_{ij}^2). \tag{9}$$

The essence of handicapping is to create fair matches. Therefore, we make the assumption that all golfers have the same mean net score, i.e. $\mu_{ij} = \mu_j$. In addition, we are going to make the clearly false assumptions that $\mu_j = \mu$ and $\tau_{ij} = \tau_i$, that the mean net scores and the net score variances are the same on

all holes. However, this assumption is not problematic as the same analysis can be undertaken on a hole-by-hole basis leading to the same proposal for team compositions. Without loss of generality, we also set $\mu = 0$ as it is only comparative golf scores that are relevant. Accordingly, we simplify (9) whereby the net score for golfer i on each hole is given by

$$Y_i \sim \text{Normal}(0, \tau_i^2). \tag{10}$$

With a two-man team consisting of players i_1 and i_2 , the quantity of interest is the distribution of the net best-ball result

$$Z_{i_1, i_2} = \min(Y_{i_1}, Y_{i_2}).$$

It is shown by Nadarajah and Kotz (2008) that Z_{i_1, i_2} is nearly normal if τ_{i_1} and τ_{i_2} do not vary greatly. Using (10), assuming that τ_{i_1} and τ_{i_2} do not vary greatly and assuming the independence between Y_{i_1} and Y_{i_2} , the moment expressions (11) and (12) from Nadarajah and Kotz (2008) lead to the approximate distribution

$$Z_{i_1, i_2} \sim \text{Normal}\left(-\frac{1}{\sqrt{2\pi}}(\tau_{i_1}^2 + \tau_{i_2}^2)^{1/2}, \left(\frac{\pi - 1}{2\pi}\right)(\tau_{i_1}^2 + \tau_{i_2}^2)\right). \tag{11}$$

If we pair golfers such that every pair has the same probability distribution, then each pair has the same probability of finishing in any position in a tournament. Therefore, if the i_1 's and i_2 's are paired such that $\tau_{i_1}^2 + \tau_{i_2}^2 = c$ for some constant c , then the objective is achieved as each distribution in (11) is $\text{Normal}(-c^{1/2}/\sqrt{2\pi}, c(\pi - 1)/(2\pi))$.

Therefore, we have a prescription for pairing golfers in two-man net best-ball tournaments. We use $\hat{\sigma}$ in (8) as a proxy for τ , and we simply match the golfer with the highest $\hat{\sigma}$ with the golfer with the lowest $\hat{\sigma}$, we match the golfer with the second highest $\hat{\sigma}$ with the golfer with the second lowest $\hat{\sigma}$, and so on. Given the $\hat{\sigma}$ values, the forming of two-man teams is an easy task for the golf director.

In the case of four-man net best-ball tournaments, the investigation of team composition ought to consider the distribution of $Z_{ijkl} = \min(Y_i, Y_j, Y_k, Y_l)$. However, the distribution theory for Z_{ijkl} appears

Table 1
The weights q_i in (8) that are used in the calculation of the spread $\hat{\sigma}$

q_1	q_2	q_3	q_4	q_5	q_6	q_7	q_8
-0.1511	-0.1006	-0.0792	-0.0632	-0.0500	-0.0384	-0.0277	-0.0178
q_9	q_{10}	q_{11}	q_{12}	q_{13}	q_{14}	q_{15}	q_{16}
-0.0082	0.0011	0.0103	0.0196	0.0291	0.0389	0.0492	0.3880

intractable and we therefore propose an expedient approach. Our heuristic begins with the optimal two-man teams described above, and we then combine pairs of the two-man teams based on the mean values in (11). Our procedure ranks the two-man teams according to $\hat{\sigma}_{i_1}^2 + \hat{\sigma}_{i_2}^2$. We then match the two-man team with the highest $\hat{\sigma}_{i_1}^2 + \hat{\sigma}_{i_2}^2$ with the two-man team with the lowest $\hat{\sigma}_{i_1}^2 + \hat{\sigma}_{i_2}^2$, and so on. Again, given the $\hat{\sigma}$ values, this is a simple task for the golf director.

There may be alternative approaches for forming four-man teams. For example, one could stratify the golfers into four groups according to increasing $\hat{\sigma}_i$ and then form a four-man team by randomly selecting a golfer from each group. In the simulation study of Section 4.2, it turns out that a slight variation of our approach (which we refer to as WCS_Z) gives even better results.

3.1. The normality assumption

Whereas the normality of golf scores is frequently assumed in the literature (e.g. Pollock (1977), Scheid (1990), Berry (2010)), these papers assume normality on 18-hole scores. However, in (10), normality is assumed on individual holes. This strong assumption does not strike us as problematic for several reasons. First, the development that follows (10) does not utilize the full distributional properties of the normal. Rather, only the first two moments are used in determining team composition. Second, instead of defining Y_i as the net score of the i th golfer, we could have alternatively defined Y_i as the latent net skill or performance of the i th golfer. Net skill/performance is a continuous variable that is clearly concave and better resembles normality. Although not as clean, such a formulation would have led to the same approach for team composition. We can therefore think of golf scores as a discretized version of skill/performance. But the main point is that the normality argument is provided only as a heuristic for forming teams, and we use actual golf data and simulation to establish that the approach improves upon traditional practice. In Section 4, we supplement the theoretical underpinnings with simulation studies.

4. Simulation studies

A rationale for the proposed team composition was provided using statistical theory in Section 3. How-

ever, given that the statistical theory was based on some approximations, it is good to supplement the theory via simulation. We first generate golf scores from a theoretical model for scoring. We then use a resampling scheme to generate golf scores from a dataset of actual golf scores.

An aspect of our evaluation procedure which differs from the literature concerns our definition of fairness. The typical definition of a fair tournament involving n teams is that each team wins a tournament with probability $1/n$ (Pavlikov et al. 2014 and Benincasa et al. 2017). Imagine a pathological example involving four teams where a particular team ends up in first, second, third and fourth place 25%, 0%, 0% and 75% of the time. By strictly considering win probabilities, the competition is fair for this team. However, it is not fair in the sense that 3/4 of the time, the team will end up in last place. This would be problematic if there is prize money for first place, second place, third place and fourth place. This is a motivating example for our matrix definition of fairness where a fair tournament involving n teams is one where for each team

$$\text{Prob}(\text{finishing in } j\text{th place}) = 1/n, \quad j = 1, \dots, n.$$

4.1. Simulation via a theoretical scoring model

In Table 2, we provide probability distributions for 8 fictitious golfers corresponding to their performance on each hole. The distributions are not entirely realistic as we only permit the four net scores of birdie (-1 relative to par), par, bogey (+1 relative to par) and double-bogey (+2 relative to par). However, the probability distributions have been constructed such that each golfer has the same mean net score which is consistent with the desiderata of the handicapping system. The most noteworthy aspect of Table 2 is that the performance of the golfers is variable with increasing standard deviations as we go down the rows of the table. Therefore, golfer 1 is the most consistent and golfer 8 is the most variable.

Imagine that these 8 golfers are competing in teams of size $m = 2$. Therefore, the number of possible tournament constructions is $\binom{8}{2}\binom{6}{2}\binom{4}{2}/4! = 105$. Consider one such tournament construction. For each team, we first generate 18 holes for each golfer in the pair according to their probability distributions in Table 2. We then determine the team's aggregate net best-ball score T according to (3). For this particular round of 18 holes and for the particular tournament construction, we determine the finishing order of the

Table 2

Net score probability distributions for 8 fictitious golfers

Golfer	Net Score Probability Distribution				Mean	SD
	-1	0	1	2		
	1	0.02	0.96	0.02		
2	0.06	0.88	0.06	0.00	0.00	0.346
3	0.10	0.80	0.10	0.00	0.00	0.447
4	0.14	0.72	0.14	0.00	0.00	0.529
5	0.18	0.64	0.18	0.00	0.00	0.600
6	0.20	0.61	0.18	0.01	0.00	0.648
7	0.26	0.51	0.20	0.03	0.00	0.762
8	0.32	0.41	0.22	0.05	0.00	0.860

Table 3

Percentage of time in the WCS simulation corresponding to the four finishing positions for each of the four teams

	1&8	2&7	3&6	4&5
Finish 1st	37.4	24.7	17.7	20.2
Finish 2nd	28.5	25.4	20.6	25.5
Finish 3rd	21.5	27.0	27.8	23.7
Finish 4th	12.6	22.9	33.9	30.6

four teams. We repeat the simulation procedure for 1,000 tournaments to obtain frequency tables for the finishing positions. Note that if a tie exists for a particular round, we randomly break ties.

According to our proposed team composition developed in Section 3, the optimal tournament construction in terms of fairness is 1&8, 2&7, 3&6 and 4&5. We denote this tournament construction as WCS (an acronym based on the authors' surnames). We are interested in how WCS performs compared to the other 104 potential tournament constructions. Table 3 provides the percentage of time in the WCS simulation corresponding to the four finishing positions for each of the four teams. If the competition were completely fair, the table entries would all be equal (i.e., 25.0%). For WCS, we observe that Team 1&8 is clearly the strongest team finishing in the top two positions nearly 66% of the time.

However, whether WCS is meritorious can only be determined in the context of the other potential tournament constructions. And each tournament construction has a corresponding Table 3 resulting from the simulation procedure. For each tournament construction, it is natural to assess fairness via the Chi-Square test statistic

$$\tilde{\chi}^2 = \sum_{i=1}^4 \sum_{j=1}^4 \frac{(f_{ij} - E_{ij})^2}{E_{ij}} \quad (12)$$

Table 4

The five best and five worst tournament constructions for the golfers in Table 2 ordered according to the Chi-Square test statistic (12)

Ranking	Team A	Team B	Team C	Team D	Chi-Square test statistic
1 (WCS)	1 & 8	2 & 7	3 & 6	4 & 5	221.4
2	1 & 8	2 & 7	3 & 5	4 & 6	270.4
3	1 & 7	2 & 8	3 & 6	4 & 5	371.2
4	1 & 8	2 & 6	3 & 7	4 & 5	420.8
5	1 & 7	2 & 8	3 & 5	4 & 6	433.0
...
101	1 & 2	3 & 6	4 & 5	7 & 8	3705.0
102	1 & 2	3 & 4	5 & 7	6 & 8	3738.7
103	1 & 2	3 & 4	5 & 8	6 & 7	3851.5
104	1 & 2	3 & 5	4 & 6	7 & 8	3919.3
105	1 & 2	3 & 4	5 & 6	7 & 8	4047.0

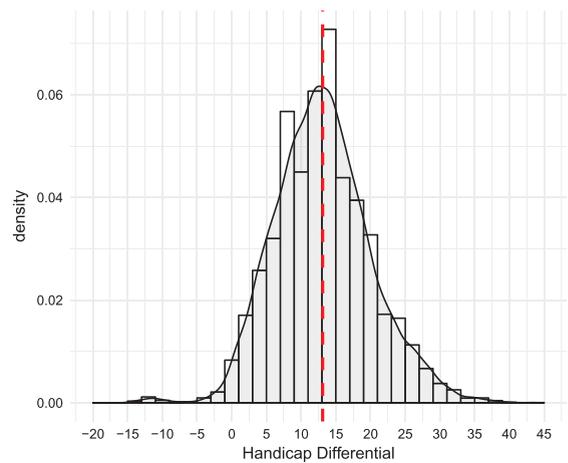


Fig. 1. Histogram and density plot of the handicap differentials for the Coloniale dataset.

where $E_{ij} = 250$ and the frequency f_{ij} is the (i, j) th entry obtained from its corresponding Table 3 matrix. Under the null hypothesis that the tournament construction is fair, $\tilde{\chi}^2$ has a Chi-Square distribution on 9 degrees of freedom. Large values of $\tilde{\chi}^2$ provide evidence against the null hypothesis. Table 4 lists the best performing and worst performing tournament constructions based on the Chi-Square test statistic (12). Although none of the team constructions are fair (i.e., they all have p-values that are statistically significant), we observe that our proposed team formation WCS is the best possible tournament construction. It is the best construction that can be achieved given the characteristics of the 8 golfers and the rules of net best-ball tournaments. We also note that the best five tournament constructions are similar in the sense that golfers with high variability are paired with ones with low variability.

Table 5
 Demonstration of the Zigzag method of team composition where 16 golfers are formed into four teams

	Golfers ordered by handicap index from the lowest to highest															
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
Team 1	x							x	x							x
Team 2		x					x			x					x	
Team 3			x			x					x			x		
Team 4				x	x							x	x			

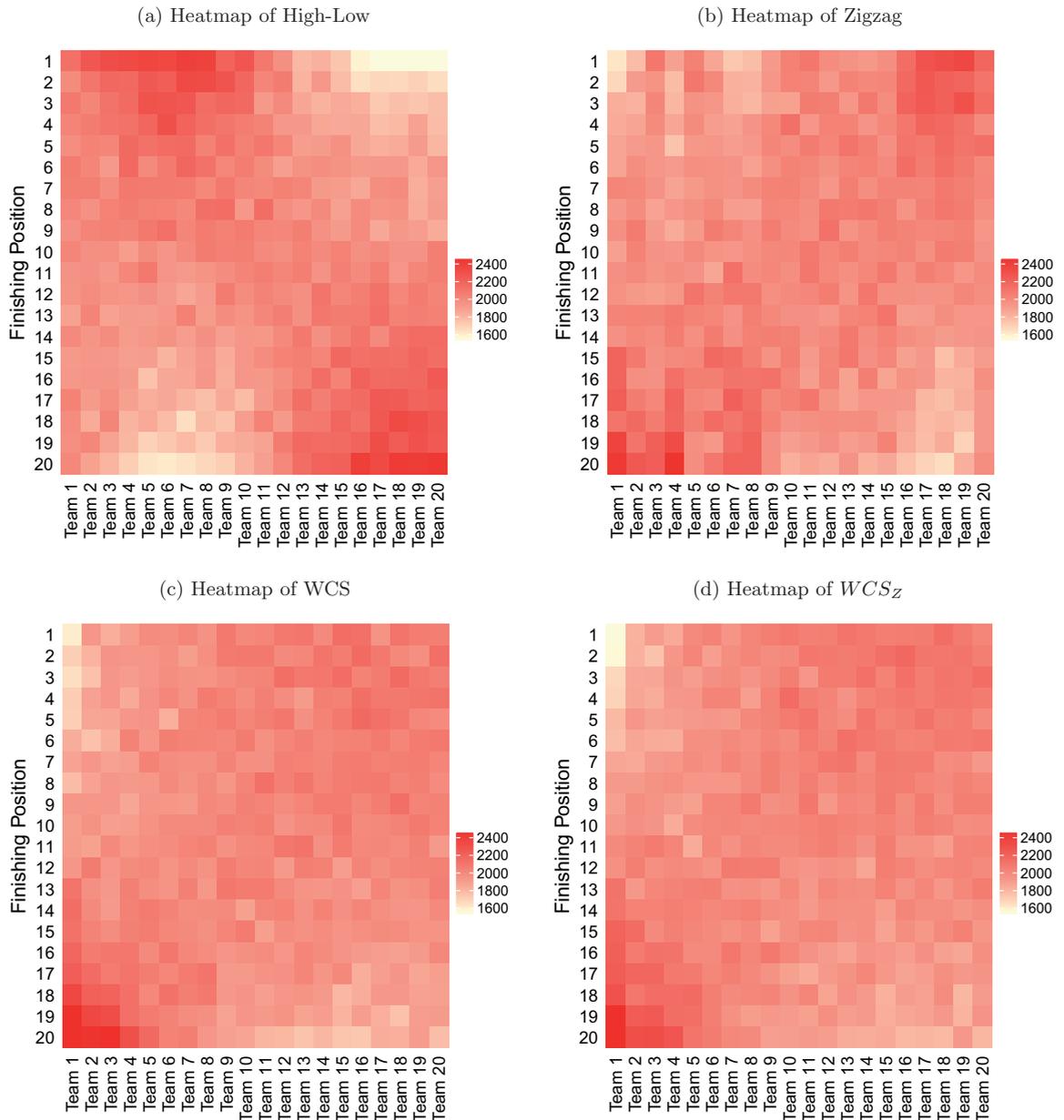


Fig. 2. Heatmaps of the frequency tables produced by High-Low, Zigzag, WCS and WCS_z .

4.2. Simulation via actual golf scores

This simulation study is based on a dataset obtained from Coloniale Golf Club in Beaumont, Alberta collected over the years 1996 through 1999.

After restricting scores to male golfers who have played at least 40 rounds, we are left with a dataset consisting of 10,470 rounds collected on 80 golfers. Therefore, the average number of rounds played per golfer in the restricted dataset is approximately 131. In Figure 1, we provide a histogram and density plot of the handicap differentials corresponding to the 10,470 rounds. The mean handicap differential is approximately 13 and is marked by the dashed vertical line. The data correspond to a large pool of golfers with varying skill levels. It should therefore have the required generality for testing our proposed method of team formation in net best-ball tournaments.

Our simulation procedure is based on a resampling scheme. In this exercise, suppose that we investigate a particular team construction heuristic where we form $n = 20$ teams of size $m = 4$. For each golfer, our first step involves randomly selecting 20 of his rounds of golf. These rounds will form his 20 differentials from which his handicap index I in (5) and his spread statistic $\hat{\sigma}$ in (8) can be calculated. These two statistics are sufficient for determining all of the common methods of team composition including our proposed method. Therefore, based on the particular team construction heuristic, the 20 teams of four players are identified. For each of these 80 golfers, we next generate one of their remaining rounds of golf for which we have hole-by-hole scores. In golf, detailed hole-by-hole data is rare and is a feature of the Coloniale dataset. Using the generated round of golf for each of the 80 golfers, each team's aggregate net best-ball score T can be calculated according to (3), and we obtain the finishing order of the 20 teams composed by the particular team construction heuristic. This resampling procedure is repeated over 40,000 hypothetical tournaments. Frequency tables are obtained as in Table 3 where finishing order corresponds to the rows and team compositions correspond to the columns. In this simulation exercise we therefore have matrices of dimension 20×20 .

We now compare two common team constructions against our proposed method WCS and a variation of WCS. Recall from Section 3 that our method of team formation first ranks golfers according to $\hat{\sigma}$, and then pairs golfers 1&80, 2&79, and so on. Then, these 40 pairs are ranked according to $\hat{\sigma}_{i_1}^2 + \hat{\sigma}_{i_2}^2$ where i_1 and i_2 are in the same pair. We then pair the pairs as

Table 6

Comparison of the four methods of team formation based on the Chi-square test statistic

Method	$\bar{\chi}^2$
High-Low	4049.09
Zigzag	2136.88
WCS	2034.99
WCS _Z	1742.38

before with high values of $\hat{\sigma}_{i_1}^2 + \hat{\sigma}_{i_2}^2$ matched with low values. This algorithm determines the 4-man teams.

We refer to the most common method of team formation as “High-Low” where High-Low is very similar in construction to WCS. The only difference is that orderings are based on the handicap index I in (5) rather than spread statistic $\hat{\sigma}$ in (8). The High-Low heuristic is that strong golfers are matched with weak golfers in the first pairing, and then strong teams (based on cumulative handicap indices) are matched with weak teams in the subsequent pairing.

We refer to the third method of team formation as “Zigzag” which is less common than High-Low. Pavlikov, Hearn and Uryasev (2014) provide an illustrative example of Zigzag. For simplicity, consider the formation of 16 golfers into four teams of four players as shown in Table 5. For example, Team 1 consists of golfers 1, 8, 9 and 16. The intuition behind Zigzag is that the summation of handicap indices should be nearly constant across teams.

The fourth method which we refer to as “WCS_Z” is a “Zigzag” variation of WCS. In WCS_Z, we carry out the Zigzag team formation procedure but instead of using handicap indices, we use the spread statistic $\hat{\sigma}$.

In Table 6, we provide the results of the comparison using the simulated frequency tables based on High-Low, Zigzag, WCS and WCS_Z. In this exercise, the Chi-Square statistic (12) is used to assess the four methods of team formation where the summations in (12) extend over the 20×20 cells and $E_{ij} = 2,000$. We observe that WCS_Z outperforms the other three methods in terms of giving the lowest Chi-Square statistic. The High-Low method is clearly the worst of the three methods in terms of fairness.

Another way of assessing the four methods is via heatmaps. In Figure 2, we produce heatmaps corresponding to the simulated frequency tables based on High-Low, Zigzag, WCS and WCS_Z. It is evident that the new approaches (WCS and WCS_Z) have more constant coloring; this indicates that they are fairer

methods of team composition than both High-Low and Zigzag.

5. Discussion

In most sports, it is only reasonable for players of comparable abilities to compete. For example, it is difficult to imagine any basketball related competition where an average person is matched up against LeBron James.

However, in golf, a comprehensive handicap system has been devised to allow players of different abilities to compete fairly against one another. Unfortunately, the handicap system can be far from fair in particular competitions, and there are many types of competitions in golf. For example, golf can be played according to match play or stroke play, golf can be played 1v1, 2v2 or in tournament settings, and golf can be played in various formats such as best-ball, foursomes, aggregate, scrambles, etc.

In this paper, we have devised simple proposals where teams of sizes two and four are formed in net best-ball competitions. Using both statistical theory and simulation studies, we have demonstrated that the proposals are more fair than standard procedures for team composition.

One golfing format which is rarely used is a “worst-ball” competition where in two-man teams (for example), the higher of the two scores is the recorded team score. Following the development of the distribution of the minimum in (11) and using the moment expressions in (9) and (10) from Nadarajah and Kotz (2008), the distribution of the maximum is approximately Normal $\left(\frac{1}{\sqrt{2\pi}}(\tau_{i_1}^2 + \tau_{i_2}^2)^{1/2}, \left(\frac{\pi-1}{2\pi}\right)(\tau_{i_1}^2 + \tau_{i_2}^2)\right)$. Therefore, as previously argued, the prescription for optimal pairing golfers would be the same as before where the most variable golfer is matched with the least variable golfer, and so on.

The key component of our proposal is the recognition of variability in golf performance. Perhaps the variability aspect can be introduced to improve fairness in other types of golf competitions.

References

Benincasa, G., Pavlikov, K. and Hearn, D., 2017, Algorithms and software for the golf director problem. Available at www.optimization-online.org/DB_FILE/2017/10/6299.pdf (Accessed: 23 July 2018).

- Berry, S.M., 2010, Is Tiger Woods a winner? In *Mathematics and Sports*, Dolciani Mathematical Expositions #43, J.A. Gallian, editor, Mathematics Association of America: Washington, 157-168.
- Bingham, D.R. and Swartz, T.B., 2000, Equitable handicapping in golf, *The American Statistician*, 54(3), 170-177.
- Chan, T., Madras, D. and Puterman, M., 2018, Improving fairness in match play golf through enhanced handicap allocation, *Journal of Sports Analytics*, Prepress, 1-12.
- Golf Canada, 2016, Golf Canada Handicap Manual. Available at <http://golfcanada.ca/app/uploads/2016/02/2016-Handicap-Manual-6x9-ENG-FA-web.pdf> (Accessed: 21 August 2017).
- Grasman, S.E. and Thomas, B.W., 2013, Scrambled experts: Team handicaps and win probabilities for golf scrambles, *Journal of Quantitative Analysis in Sports*, 9, 217-227.
- Hurley, W.J. and Sauerbrei, T., 2015, Handicapping net best-ball team matches in golf, *Chance*, 28, 26-30.
- Kupper, L.L., Hearne L.B., Martin, S.L. and Griffin, J.M., 2001, Is the USGA golf handicap system equitable? *Chance*, 14(1), 30-35.
- Nadarajah, S. and Kotz, S., 2008, Exact distribution of the max/min of two gaussian random variables, *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*, 16(2), 210-212.
- Pavlikov, K., Hearn, D. and Uryasev, S., 2014, The golf director problem: Forming teams for club golf competitions. In *Social Networks and the Economics of Sports*, P.M. Pardalos, and V. Zamaraev, editors, Springer International Publishing: Switzerland, 157-170.
- Pollock, S.M. 1974, A model for the evaluation of golf handicapping, *Operations Research*, 22(5), 1040-1050.
- Pollock, S.M. 1977, A model of the USGA handicap system and fairness of medal and match play. In *Optimal Strategies in Sports*, S.P. Ladany and R.E. Machol, editors, North Holland: Amsterdam, 141-150.
- Scheid, F.J. 1977, An evaluation of the handicap system of the United States Golf Association. In *Optimal Strategies in Sports*, S.P. Ladany and R.E. Machol, editors, North Holland: Amsterdam, 151-155.
- Scheid, F.J. 1990, On the normality and independence of golf scores, with various applications. In *Proceedings of the First World Scientific Congress of Golf*, A.J. Cochran editor, E & FN Spon: London, 147-152.
- Siegbahn, P. and Hearn, D., 2010, A study of fairness in fourball golf competition. In *Optimal Strategies in Sports Economics and Management*, S. Butenko, J. Gil-Lafuente and P.M. Pardalos, editors, Springer-Verlag: Heidelberg, 143-170.
- Swartz, T.B., 2009, A new handicapping system for golf, *Journal of Quantitative Analysis in Sports*, 5(2), Article 9.
- USGA 2016, USGA Handicap System Manual. Available at <http://www.usga.org/Handicapping/handicap-manual.html#!rule-14367> (Accessed: 21 August 2017).
- Yun, H., 2011a, History of handicapping, part I: Roots of the system. Available at <http://www.usga.org/articles/2011/10/history-of-handicapping-part-i-roots-of-the-system-21474843620.html> (Accessed: 21 August 2017).
- Yun, H., 2011b, History of handicapping, part II: Increasing demand. Available at <http://www.usga.org/content/usga/>

home-page/articles/2011/10/history-of-handicapping-part-ii-increasing-demand-21474843658.html (Accessed: 21 August 2017).

Yun, H., 2011c, History of handicapping, part III: USGA leads the way. Available at <http://www.usga.org/content/usga/home-page/articles/2011/10/history-of-handicapping-part-iii-usga-leads-the-way-21474843686.html> (Accessed: 21 August 2017).

Yun, H., 2011d, History of handicapping, part IV: The rise of the slope system. Available at <http://www.usga.org/content/usga/home-page/articles/2011/10/history-of-handicapping-part-iv-the-rise-of-the-slope-system-21474843751.html> (Accessed: 21 August 2017).