WORKING PAPER 13/2022 (ECONOMICS AND STATISTICS)

Are Some Athletes More Cognitive Skilled than Others when Choosing their Opponents in Skiing-Sprint Elimination Tournaments?

Niklas Karlsson and Anders Lunander

ISSN 1403-0586

Örebro University School of Business
SE-701 82 Örebro, Sweden

Are Some Athletes More Cognitive Skilled than Others when Choosing their Opponents in Skiing-Sprint Elimination Tournaments?

Niklas Karlsson*

School of Business, Örebro University, Sweden

Anders Lunander**

School of Business, Örebro University, Sweden

November 2022

Abstract

The study analyzes data from world cup cross-country skiing sprint elimination tournaments for men and women in 2015-2020. In these tournaments prequalified athletes sequentially choose in which of five quarterfinal heats they want to compete. Due to a time constraint on the day the tournament is held, the recovery time between the elimination heats varies. This implies a clear advantage for the athlete to race in an early rather than in a late quarterfinal to increase the chances of being successful in a possible final. Given that athletes seek to maximize their expected achieved world cup points when choosing quarterfinal, a simple model predicts that higher ranked athletes prefer to compete in early quarterfinals, despite facing expected harder competition. The result is consistent with our empirical analysis of the data. We also develop two estimation methods to investigate whether some athletes are found to be more tactical skilled in their decision making. Our estimates indicate that twelve out of 115 athletes have made choices having an expected positive effect on their performance in terms of achieved world cup points. For 22 athletes the effect is expected to be negative. The estimated individual effects ranges from -3 points to +4 points.

JEL Classifications: C51, C72, D91, Z20

Keywords: elimination tournament, game theory, choosing opponents, skiing sprint, tactical skills

* Örebro University, School of Business, 701 82 Örebro, Sweden
e-mail: niklas.karlsson@oru.se
** Örebro University, School of Business, 701 82 Örebro, Sweden
e-mail: anders.lunander@oru.se (corresponding author)

1 Introduction

A salient feature in the design of the elimination tournaments in cross-country skiing sprint for men and women, organized by the International Ski Federation (FIS), is that prequalified athletes choose themselves in which of five quarterfinals they want to compete in. This design was implemented for the first time at the end of the ski season 2014/2015 and the mechanism replaced a standard seeding method which was applied to equalize expected competition across quarterfinals, i.e., a balanced seeding. The 30 fastest athletes from an initial qualification round now choose sequentially their quarterfinal, six athletes in each quarterfinal. The order of the sequential choices is based on the ranking from the qualification round.[1] Each athlete knows the previous choices of the others before making her own choice of quarterfinal. The distribution of athletes across the subsequent races, the two semifinals and the final – each race containing six contestants -, then follows a known scheme. In Karlsson and Lunander (2020) (KL henceforth) a background to the FIS's motives to replace the quarterfinal seeding scheme for a new design is provided, as well as a detailed description how these competitions are carried out. The authors recognize the complex optimization problem the athletes face when choosing their optimal quarterfinals. In a simple model, they seek to capture the athlete's strategic choice of quarterfinal when facing a trade-off between recovery time and expected competition. On one hand, the advancement from an early race to the next elimination round means that the athlete can benefit from a longer recovery time before the next round, than those athletes advancing from a late race can do. On the other hand, the competition in an early race is expected to be tougher than it is expected to be in a late race, thus making it harder for the athlete to qualify for the next round when choosing an early race. Using data from cross-country skiing sprint competitions 2015-2020 for men and women, they find that their modelled prediction – a higher ranked athlete from the qualification round tends to choose an early quarterfinal rather than a late quarterfinal – is consistent with the athletes' observed choices of quarterfinals. Their econometric analysis suggests that higher ranked athletes even underestimate the value of choosing an early quarterfinal. The probability of reaching the podium is still higher when choosing an early quarterfinal when conditioning on athletes' capacity.

However, the modelled behavior in KL rests on the assumption that the athlete maximizes her probability of winning the final. They apply a logistic regression model to estimate the athlete's probability of reaching the podium as a function of her choice of quarterfinal, her individual capacity relative other athletes, and individual specific effects.

Our study has two main purposes. The purpose of the first part of the study is to extend the work in KL by assuming another objective function. Our empirical analysis is based on the same data used in KL, that is, official results from 34 competitions in seasons 2014/2015 -2019/2020 and information about the 30 athletes competing in the quarterfinals, men as well as women. Our point of departure is that the athlete, instead of maximizing her probability of winning the tournament, maximizes her expected award of world cup points when choosing her quarterfinal, which is likely to be a plausible objective function for the majority of the 30 individuals from the qualification round. To gain theoretical understanding of the problem, we first adopt the simple elimination tournament model presented in KL, with two rounds – two semifinals and one final - and four players, replacing the model's objective function. Applying the statistical test derived in KL, we then test the modelled prediction on the data. Finally, we carry out an OLS regression analysis to examine what effect the athletes' choices of quarterfinals has had upon their achieved world cup points. The results we obtain are all in line with the results presented in KL.

The second part contains a novel analysis on cognitive ability and performance in sports. The purpose of this analysis is to examine to what extent athletes have been able to take advantage of the cognitive challenge that followed from the replacement of a balanced seeding scheme of the five quarterfinals with a choice-of-quarterfinal mechanism. Given the athletes' physiological capacities and observed

[1] Given the ranking of the 30 qualified athletes from the qualification round, the choice of quarterfinal is first done in descending ranking order, starting with the athlete having ranking number 11, followed by number 10, number 9 and so on. As soon as the athlete with ranking number 1, that is, the fastest athlete from the qualification round, has made her choice, the remaining nineteen athletes make their choices in ascending ranking order, starting with the athlete having ranking number 12 {12, 13,….29, 30}.

performance, we investigate whether any of them are found to be more cognitive skilled than others at making these demanding decisions. We develop an estimation approach that indicates how well athletes in their choices of quarterfinals have managed to balance the trade-off between recovery time and level of competition.

Our paper is organized as follows: the skiing sprint competitions and the procedure to choose quarterfinals are presented in Section 2. Section 3 provides a literature overview. The model is given in Section 4, followed by a description of the data in Section 5. The empirical methodology and the outcome of our tests are presented in Sections 6-8 and, finally Section 9 presents the conclusions.

2 The Skiing Sprint Competition

A FIS's skiing sprint competition takes place on one day with men and women competing in separate classes. The competition is an elimination tournament, which is preceded by a qualification round. In the qualification round, each of about 60 to 80 athletes ski a course of about 1.5 km length, starting at about 15 second intervals. The 30 fastest athletes qualify for the five quarterfinals, with six athletes in each quarterfinal. The five quarterfinals are then followed by two semifinals and a final, each race with six athletes. The top-two athletes in each quarterfinal are directly qualified for the semifinals, where the athletes coming from the first two quarterfinals are placed in the first semifinal, and the athletes coming from the two last quarterfinals are placed in the second semifinal. The top-two athletes from the third quarterfinal are placed in different semifinals, putting the athlete in first place in the first semifinal. In addition to these top-ten athletes from the five quarterfinals, the two athletes with the best times of the athletes ending up at place 3-6 in the five quarterfinals (the lucky losers) are also qualified for the semifinals. The faster of these two is placed in semifinal one while the other athlete is placed in semifinal two. The elimination races are run on the same course as the qualification round and mass start is applied. The distribution of the athletes to the five quarter-finals does not follow a predetermined scheme, but the athletes themselves choose sequentially in which quarter-final they want to go. Given the ranking of the 30 qualified athletes from the qualification round, the choice of quarterfinal is first done in descending ranking order, starting with the athlete having ranking number 11, followed by number 10, number 9 and so on. As the athlete with ranking number 1, that is, the fastest athlete from the qualification round, has made her choice, the remaining nineteen athletes make their choices in ascending ranking order, starting with the athlete having ranking number 12 {12, 13,….29, 30}. The switch of mechanism, from a balanced seeding mechanism to a "choosing procedure", to assign athletes quarterfinals, was undertaken because the result lists from the competitions held before 2014/2015 indicated that athletes, being seeded into one of the two first quarterfinals, had a much greater chance to reach the podium in the final than those athletes having raced in the two last quarterfinal. Due to the timing of the various knock-out races during the day of the competition, an athlete, advancing to the final from the first semifinal – which mainly consists of athletes having advanced from the two first quarterfinals - can benefit from a longer recovery time prior to the final than she can when advancing from the second semifinal to the final.

3 Literature

The first part of our work in this study follows the analysis provided in Karlsson and Lunander (2020), the main difference being that we make use of another objective function when modelling the athlete's optimal choice of opponent. As shown by their review of literature, the number of studies analyzing behavior and outcomes in elimination tournaments, where athletes or teams choose their opponents instead of being seeded, is very scarce. Although there in various media seems to be a handful of reported cases in sports where this type of mechanism has been implemented, there exists to our best knowledge no scientific analysis besides Karlsson and Lunander (2020) examining the properties of a "choosing-your-opponent" mechanism in knock-out tournaments. In our second part of our work, we estimate the separate effect of the individual athlete's choice of quarterfinal in terms of her achieved world cup points. Besides our measures of an athlete's physiological capacity, the estimated effect of the choice of quarterfinal would capture the athlete's tactical ability, that is, a measure of her cognitive skills in making an optimal decision. The correlation between perceptual-cognitive skills and athletes' performance in sports has been analyzed in a large body of literature. Different approaches to measure perception and cognitive functioning of experts and athletes can be found in this research. One approach focuses on the examination of parameters reflecting the interaction between the athlete and her specific environment (e.g eye movements, gaze behavior, response time, visual fixation). Mann et al. (2007) conduct a meta-analysis, covering 42 studies and find that expert athletes are more accurate and quicker in their responses than are the nonexpert athletes. Another research approach investigates the relation between performance in sports and athletes' performance in standardized cognitive laboratory tests, assumed to be relevant in their competitive sport training. In a meta-analysis of 20 studies, Voss et al. (2010) show that athletes perform better than do nonathletes. A third approach analyzes the cognitive skills between high-performance level athletes and low-performance level athletes, without linking the laboratory cognition tests to the athletes' competitive training. In their meta-analysis of 19 studies, Scharfen and Memmert (2019) find that the former group have superior cognitive functions compared with the latter group. Also, in a larger meta-analysis, examining 142 studies on cognition and sports performance, Kalén et al. (2021), find support for that higher skilled athletes perform better on cognitive function tests than lower skilled athletes.

The contribution of our study to the existing literature is we can observe and estimate what effect an athlete's strategic choices in a row of identical competitions have upon her outcome. The choice of quarterfinal in skiing-sprint competitions adds a cognitive dimension to the competitions, where athletes with superior cognitive functions are favored. Unlike most studies on cognition and sports performance, our study, roughly speaking, takes the laboratory experiment on cognition to the field and links an athlete's ability to make rational choices to her performed results.

4 A Model for the Process of Rational Choice

In this section we follow the head-to-head knockout tournament model provided in Karlsson and Lunander (2020). Four players, $(i = A, B, C, D)$ of two capacity types - high capacity (H) and low capacity (L) - compete head-to-head in two semifinals (s.1 and s.2), where the winners make up for the victory in a subsequent final. Players A and B are assumed to be of type H whereas players C and D are of type L. Unlike their model – in which a player maximized her probability of reaching the podium – we assume a player's objective to be to maximize her expected achieved world cup points, given her type of capacity and recovery duration between the semifinal and the final. The winner gets k_1 points and the player ending up second gets k_2 points, where $0 < k = \frac{k_2}{k_1} < 1$. Since the result of the players choice does not depend on k_1 and k_2 for a given k, without loss of generalizability we set $k_1 = 1$. The players choose in sequential order, starting with player A, which one of the two semifinals to compete in.[2] A player's choice of semifinal becomes public prior to the next player's choice. Hence, there are six possible settings $(j = 1, \ldots, 6)$ of semifinals, including the mirroring of identical plays, albeit in different order. Figure 1 illustrates the decision tree of our model.

[2] Player D never makes a choice and player C can only choose her opponent when player B chooses the opposite semifinal as player A does.

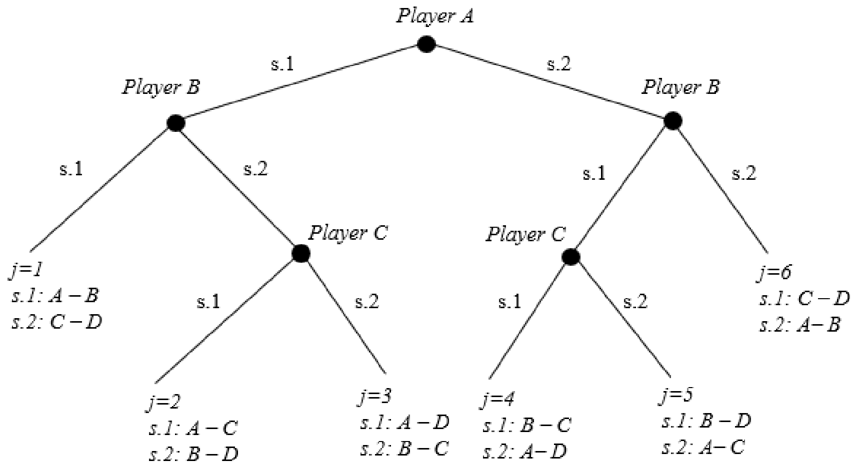

Figure 1. The sequential order of choices and possible settings of semifinals

We denote the player *i*:th probability of ending up on *r*:th place in the tournament given, setting *j*, as $p_{i,j}^{(r)}$. Furthermore, $E_{i,j}$ denotes the expected achieved world cup points for the player. For example, $p_{C,4}^{(1)}$ denotes the probability that player C will win the tournament given that she faces player B in semifinal 1, while $E_{B,2}$ denotes the expected achieved world cup points for player B facing D in semifinal 2.

A player of type *H* will beat a player of type *L* with probability *p > 0.5*, and if two players of the same type compete, the probability is 0.5 to win against the other. To capture the effect of having a shorter recovery time when the player advances into the final from the second semifinal, we multiply that player's probability of winning the final with a constant *c*, where $0 < c < 1$. The lower value of *c*, the larger is the negative effect of having advanced to the final from the second semifinal rather than from the first semifinal. Thus, even though semifinal settings *j=1* and *j=6* imply identical plays, we have $p_{A,1}^{(1)} = p_{B,1}^{(1)} > p_{A,6}^{(1)} = p_{B,6}^{(1)}$ and $p_{C,1}^{(1)} = p_{D,1}^{(1)} < p_{C,6}^{(1)} = p_{D,6}^{(1)}$ due to the recovery effect captured by *c*. The players are assumed to have full information on the values of the probabilities defined above, as well as the value of *c* and the points k_1 and k_2.

Proposition 1

Player A will always choose the first semifinal.

Player B will choose the first semifinal if and only if

$$c < f(p,k) = \frac{0.5 - pk}{(0.5 - 0.5p + 1.5p^2 - p^3)(1-k)}.$$

Otherwise, player B will play against player C in the first semifinal.

Proof: See appendix.

In Figure 2 we illustrate the result graphically. Points below the convex graph for a certain value of k, here exemplified by $k = 0, 0.2, 0.5$, indicate combinations of levels of the recovery effect and the probability *p* for which player B chooses to compete in the first semifinal.

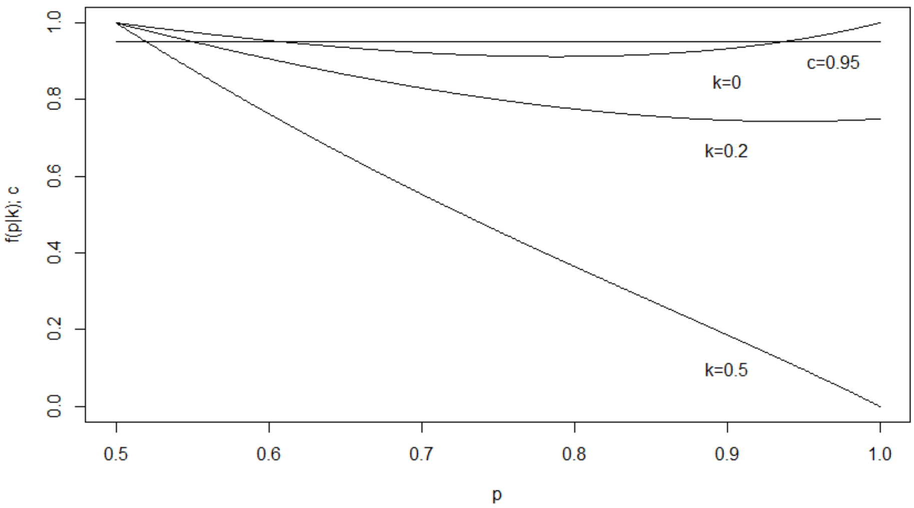

Figure 2. Illustration of player B's choice of semi-final for various combinations of p, c and k

Starting with the special case when the winner takes it all, i.e., $k = 0$, for values of c up to about 0.91, the degree of competition from low-capacity players has no effect upon player B's choice. Player B might choose to avoid player A in the first semifinal for higher values of c. For example, for $c = 0.95$, shown as a horizontal line in the figure, player B will choose the second semifinal for values of p in between 0.61 and 0.94, rounded to two decimal places.

Continuing with the case when the winner takes it all, to understand the mechanism behind player B's choice for high values of c once player A has chosen the first semifinal, we initially assume that $c = 1$ and $p = 0.5$. Thus, no recovery effect is assumed, and competition is equalized across all players. This combination of c and p obviously makes player B indifferent between the two semifinals. Now, letting p increase, still assuming no recovery effect, player B will choose the second semifinal to have a positive probability of avoiding player A in the tournament, now the single most competitive opponent. Thus, as p increases, for player B to be indifferent, a compensation in terms of a decrease in c is necessary, and we are moving downwards along the graph of the convex function in the figure. As p further increases, holding c fixed, there is a positive effect on the incentive for player B to choose the second semifinal. The importance of avoiding player A in the tournament increases due to a decrease of the relative competitiveness of type L players. However, there is also a negative effect, since the probability increases for player B of ending up weakened in the final against player A. For p about 0.79 these opposite effects cancel each other out. This negative effect outweighs the positive effect for larger values of p, meaning that we are moving upwards along the graph for indifference. For a further increase in p, player B needs to be compensated by an increase in c, i.e., a lower recovery effect, to still be indifferent and not choosing the first semifinal. For values of p close to 1, player B expects to face player A in the final with almost certainty, both advancing from different semifinals. Player B is then better off playing against player A already in the first semifinal, unless the recovery effect is negligible, i.e., c is close to *1*, making player B indifferent.

Moreover, there is also an effect of changing k. As k increases, holding c fixed, the range of values for p resulting in player B is facing player A in the first semifinal decreases. In fact, it can be shown that $\frac{\partial f(p,k)}{\partial k} < 0$. The intuition for that should be clear. As k increases, the value of merely reaching the final versus winning the tournament increases. Therefore, a high probability of reaching the final becomes more important, which is accomplished for player B by avoiding player A in the semifinal.

Finally, a prediction from the model is that the proportion of high-capacity players in the first semifinal is never low. Depending on the parameters c, p and k, either one or two high-capacity players will be found in the first semifinal, a prediction we seek support for in the data on skiing sprint competitions, to be presented next.

5 Data

Official results for all world cup skiing sprint competitions as well as individual characteristics on athletes are collected from the International Ski Federation's website (FIS Ski, 2020).

Results from 34 competitions in seasons 2014/2015 – 2019/2020 are included in the data. The data contains information about the 30 athletes competing in the quarterfinals in each competition, men as well as women. The athletes' choice of quarterfinals, their achieved results, and their official ranking are observed. The data set contains 2040 observations, divided into men and women equally.

The statistics provided in Table 1, indicate that high-capacity athletes are overrepresented in early quarterfinals, suggesting stronger competitions in the first two quarterfinals. Table 1 shows that the faster athletes from the qualification round, i.e. athletes with low qualification ranks – to greater extent choose the two first quarterfinals than the two later quarterfinals. Also, when considering the athletes' accumulated achieved world cup points prior to their choice of quarterfinal, the pattern is similar, albeit weaker for women: stronger athletes with lower accumulated ranks tend to choose early quarterfinals.

Table 1. Average qualification rank sums and average rank sums based on accumulated world cup points (in brackets) for different quarterfinals

	Quarterfinal				
Sex	*1*	*2*	*3*	*4*	*5*
Women	88.4 (92.5)	88.6 (92.1)	95.7 (93.5)	92.9 (93.1)	99.5 (93.8)
Men	87.6 (81.6)	87.7 (93.5)	93.0 (89.9)	92.6 (99.5)	104.1 (100.5)
Total	88.0 (87.0)	88.1 (92.8)	94.3 (91.7)	92.8 (96.3)	101.8 (97.2)

Moreover, as shown in Table 2, the outcome of racing in an early quarterfinal in terms of achieved world cup points is higher than racing in a late quarterfinal, despite the higher degree of competition in earlier quarterfinals. Early quarterfinalists get on average a larger amount of world cup points than late quarterfinalists.

Table 2. Average achieved world cup points for different quarterfinals

	Quarterfinal				
Sex	1	2	3	4	5
Women	28.6	27.2	22.4	21.5	20.0
Men	30.7	26.7	23.6	20.2	18.6
Total	29.7	26.9	23.0	20.8	19.3

6 Testing for Random Choice

In this section we test the null hypothesis that the athletes choose quarterfinals in a pure random way against the alternative hypothesis that athletes with high capacity to a large extent choose early quarterfinals rather than late quarterfinals, as indicated by the model from section 4. Section 6.1 presents the test statistic to be used, while the result from the test is shown in section 6.2.

6.1 The Test Statistic

We propose rank from the qualification round to capture capacity in the final rounds as well as rank among the 30 quarterfinalists based on season accumulated world cup sprint points at the time of competition. The better rank, i.e., the lower rank, the higher capacity. Thus, the prediction from the model in section 4 suggests that early quarterfinals should be overrepresented by better ranked athletes, while the opposite holds for late quarterfinals, implying that the rank sum for early quarterfinals is to undercut the rank sum for late quarterfinals.

Now, define U_k and V_k, as the rank sum for the early quarterfinals, one and two, and late quarterfinals, four and five, respectively, for competition k, $k = 1, 2, \dots, 34$. Possible values of u_k and v_k are $42, \dots, 330$. An index for type of rank sum, i.e., a rank sum based on rank from qualification or rank from season accumulated world cup sprint points, is left out to simplify notation. As is an index for sex. We formulate the null hypothesis and the alternative hypothesis as

$$H_0: E(U_k) - E(V_k) = 0 \text{ against } H_A: E(U_k) - E(V_k) < 0,$$

where the formulation of the null hypothesis is an implication of pure random choice. Following Karlsson and Lunander (2020), and adjusting for a different number of observations, it is shown that

$$\bar{R}_E - \bar{R}_L \overset{appr}{\sim} N\left(0, \frac{744}{17}\right)$$

under the null hypothesis, where $\bar{R}_E = \frac{\sum_{k=1}^{56} U_k}{34}$ and $\bar{R}_L = \frac{\sum_{k=1}^{56} V_k}{34}$ are the average rank sums and where $V(\bar{R}_E - \bar{R}_L) = \frac{744}{17}$ has been derived under consideration of the dependency between $\bar{R}_E$ and $\bar{R}_L$. Thus, an appropriate test statistic is given by

$$Z = \frac{\bar{R}_E - \bar{R}_L}{\sqrt{\frac{744}{17}}},$$

which follows approximately a standard normal distribution under the null hypothesis. Moreover, since we reject H_0 in favor of H_A for large negative values of z, the rejection region is $RR = \{z < -z_\alpha\}$where z_α is such that $P(Z > z_\alpha) = \alpha$.

6.2 Results

Consider Table 4, showing results from testing for random choice, where the figures from the first two rows are calculated from the figures in Tables 1-2. For all four possible cases, the difference in rank sums is negative as expected from the prediction from our theoretical model. However, for the case of women and rank sums based on accumulated world cup points, the negative effect is not significant, unlike the rest of the cases.

Table 4. Results from testing of random choice

	Qualification		Accumulated world cup points	
	Women	Men	Women	Men
$\bar{R}_E$	177.0	175.3	184.6	175.1
$\bar{R}_L$	192.4	196.7	186.9	200.0
$\bar{R}_E - \bar{R}_L$	-15.4	-21.4	-2.3	-24.9
z	-2.33	-3.23	-0.35	-3.76
p-value	0.0099	0.0006	0.363	0.0001

7 Modelling Achieved World Cup Points – a Regression Approach

In the previous section, statistical evidence was found that better ranked athletes were overrepresented in early quarterfinals, due to a presumed recovery effect. To estimate the recovery effect, athlete capacity and degree of competition need to be held fixed. Therefore, in this section we specify and estimate a regression model for achieved world cup points. It is modelled as a function of type of quarterfinal and various other explanatory variables capturing athlete capacity as well as degree of competition in current quarterfinals.

7.1 Model Specification

We adopt a multiple regression approach to model the k:th athlete's achieved world cup points in the skiing sprint competition l, WCP_{kl}, where $k = 1, \ldots, 30$ and $l = 1, \ldots, 68$, where $l = 1, \ldots, 34$ refers to the men's competitions and $l = 35, \ldots, 68$ refers to the women's competitions. The athlete's achieved world cup points in a competition are assumed to be conditioned on choice of quarterfinal, individual capacity relative to other athletes and degree of competition in the quarterfinal. Here follows the list of explanatory variables. The subscripts l and k are left out for simplification.

QE, dummy variable taking the value one if the athlete chooses an early quarterfinal, i.e., quarterfinals one or two, zero otherwise.

QL, dummy variable taking the value one if the athlete goes in a late quarterfinal, i.e., quarterfinals four or five, zero otherwise.

$Rqual$, rank from the qualification round minus one.

$Rwcp$, rank among the 30 quarterfinalists based on season accumulated world cup sprint points at the time of competition minus one.

$Rqual \times Rwcp$, interaction variable between $Rankqual$ and $Rankacwcp$.

$Rqualsq$, $Rqual$ squared.

$Rwcpsq$, $Rwcp$ squared.

$Rqual \times QE$, interaction variable between $Rqual$ and QE.

$Rqual \times QL$, interaction variable between $Rqual$ and QL.

$Rwcp \times QE$, interaction variable between $Rwcp$ and QE.

$Rwcp \times QL$, interaction variable between $Rwcp$ and QL.

$SRqual$, the qualification rank sum for the six athletes in the chosen quarterfinal minus 93.

$SRwcp$, the sum of ranks of accumulated world cup points for the six athletes in the chosen quarterfinal minus 93.

Thus, the multiple linear regression model can be written as[3]

$$WCP = \beta_1 + \beta_2 QE + \beta_3 QL + \beta_4 Rqual + \beta_5 Rwcp + \beta_6 Rqual \times Rwcp$$
$$+\beta_7 Rqualsq + \beta_8 Rwcpsq + \beta_9 Rqual \times QE + \beta_{10} Rqual \times QL$$
$$+\beta_{11} Rwcp \times QE + \beta_{12} Rwcp \times QL + \beta_{13} SRqual + \beta_{14} SRwcp + u,$$

where u is an error term and quarterfinal 3 is used as the reference type of quarterfinal.

[3] To allow for different effects for women and men an extended model was considered as well, also including the dummy variable *Sex* and interactions with this variable and all other explanatory variables in the model. Here, the variable *Sex* takes the value 1 if man, 0 otherwise.

The parameters are estimated using the method of ordinary least squares based on a total of 2040 observations (30 qualified athletes in each of 68 competitions). Robust standard errors are used to allow for heteroscedasticity. All computational work is performed using the program R, version 4.0.3.

We have a predetermined perception about the signs of the parameters used in the above specification.

First, the way the variables are defined, allows for an interpretation of the intercept β_1. It can be interpreted as the expected achieved world cup points for an athlete going in quarterfinal three with average degree of competition, i.e., $SRqual = SRwcp = 0$, being the winner of the qualification round ($Rqual = 0$) and at the same time is number one among the 30 quarterfinalists in the world cup standings ($Rwcp = 0$).

Second, turning to the variables representing the athlete's capacity, rank from the qualification round as well as rank based on accumulated world cup sprint points, are expected to have a negative declining effect on the expected world cup points. (Recall that low values of $Rqual$ and $Rwcp$ represent high capacity.) Thus, β_4 and β_5 are both expected to be negative, while β_7 and β_8 are expected to be positive. Furthermore, β_6 is expected to be positive. It means that we assume the negative effect of $Rqual$ on the expected WCP to diminish with the value of $Rwcp$. To put it another way, the two negative effects of $Rqual$ and $Rwcp$ are expected to reinforce each other.

Third, when it comes to the parameters related to the choice of type of quarterfinal, we note that β_2 can be interpreted as the expected difference in world cup points between racing in an early quarterfinal compared to quarterfinal three. We control for competition in terms of $SRqual$ and $SRwcp$, and at the same time we condition on an athlete who wins the qualification round and who is ranked number one among the 30 athletes in the skiing sprint world cup standings. The parameter β_3 can be interpreted analogously. Due to the recovery effect, β_2 is therefore expected to be positive, while β_3 is expected to be negative.

All athletes, irrespective of capacity, prefer an early quarterfinal to the third quarterfinal, controlling for competition, because of the recovery effect. However, the lower capacity the smaller effect on achieved world cup points, since the probability of reaching the final, and thereby benefit from the recovery effect, decreases with capacity. Therefore, we expect β_9 and β_{11} to be negative. A similar reasoning motivates β_{10} and β_{12} to be positive.

Fourth, considering the variables $SRqual$ and $SRwcp$, we expect the corresponding parameters β_{13} and β_{14} both to be positive, since the degree of competition decreases with the values these variables take on.

7.2 Estimation Results for the Regression Model

Below, consider the results of estimation of the regression model set out in section 7.1.[4]

$$\widehat{WCP} = \underset{(2.89)}{66.8^{***}} + \underset{(2.97)}{6.91^{***}} QE - \underset{(2.96)}{6.05^{**}} QL - \underset{(0.229)}{2.66^{***}} Rqual - \underset{(0.247)}{2.30^{***}} Rwcp + \underset{(0.00652)}{0.0297^{***}} Rqual \times Rwcp$$

$$+ \underset{(0.00685)}{0.0376^{***}} Rqualsq + \underset{(0.00700)}{0.0391^{***}} Rwcpsq - \underset{(0.125)}{0.235^{**}} Rqual \times QE + \underset{(0.124)}{0.143} Rqual \times QL$$

$$- \underset{(0.146)}{0.0700} Rwcp \times QE + \underset{(0.134)}{0.189^{*}} Rwcp \times QL + \underset{(0.0371)}{0.160^{***}} SRqual + \underset{(0.0266)}{0.0160} SRwcp.$$

The estimate of 66.8 for the intercept is to be interpreted as the expected achieved world cup points for an athlete going in a quarterfinal 3 with average degree of competition, who is the winner of the qualification round and at the same time is number one among the 30 quarterfinalists in the world cup standings.

For the same athlete, facing the same competition in an early quarterfinal, the expected achieved world cup points are estimated to increase with 6.91, and decrease with 6.05 if facing the same competition

[4] No statistical evidence is found for the extended model, i.e., controlling for sex, to perform better compared to the simpler model. The parameter estimates corresponding to *sex* and the thirteen interaction variables are not significant as a group ($p = 0.763$).

in a late quarterfinal. These two estimates might be interpreted as the recovery effects for such an athlete since we are controlling for the degree of competition in the model.

The parameter estimates corresponding to the variables $Rqual$, $Rwcp$, $Rqual \times Rwcp$, $Rqualsq$ and $Rwcpsq$ are, except for all having expected signs, also highly significant. Thus, we have strong support for a negative declining effect on the expected world cup points of rank from the qualification round as well as of rank based on accumulated world cup points. Moreover, we also have support for the two negative effects of $Rqual$ and $Rwcp$ to reinforce each other.

The parameter estimates for the interaction variables $Rqual \times QE$, $Rqual \times QL$, $Rwcp \times QE$, and $Rwcp \times QL$ are not all individually significant. However, they are significant as a group ($p = 0.001$). Moreover, they all have the expected signs. Therefore, we keep them in the model. To get some insights in the interpretation, consider an athlete where $Rqual = Rwcp = 10$. The expected advantage in terms of expected achieved world cup points of going in an early quarterfinal compared to quarterfinal 3 is $6.91 - 0.235 \cdot 10 - 0.07 \cdot 10 = 3.86$, i.e., $6.91 - 3.86 = 3.05$ points lower than for the top athlete referred to above.

Finally, the effects of the variables $SRqual$ and $SRwcp$ are both positive, as expected. For $SRqual$, an increase by one unit, i.e., a decrease in competition, the expected achieved world cup points are estimated to increase by 0.160, conditioning on a certain athlete's capacity going in a specific quarterfinal. However, for $SRwcp$ the effect is small and not even significant. Still, we keep the variables as a group for capturing degree of competition ($p = 0.001$).

8 Rational Choice at the Group Level

The results from Section 6 indicate that high-capacity athletes choose early quarterfinals in large part, a phenomenon that is explained by the recovery effect estimated in Section 7. Thereby, in terms of expected world cup points, those relatively few high-capacity athletes choosing late quarterfinals might get fully compensated for shorter recovery time if reaching the final, thanks to weaker competition in the rounds preceding the final. Thus, the increased chance of reaching the final, and thereby getting a quite large amount of world cup points, could possibly balance the decreased chance of performing on top once in the final. We think of such a behavior as rational, i.e., when the athletes´ choice as a group of quarterfinals as a group makes the expected achieved world cup points irrespective of type of quarterfinal, when conditioning on the athlete's capacity. A reduced form of the model specified in section 7, where the variables measuring competition is dropped, will serve as a framework to test whether the athletes as a group make rational choices.

8.1 A Test for Rational Choice

Consider the model

$$WCP = \beta_1 + \beta_2 QE + \beta_3 QL + \beta_4 Rqual + \beta_5 Rwcp + \beta_6 Rqual \times Rwcp + \beta_7 Rqualsq$$
$$+ \beta_8 Rwcpsq + \beta_9 Rqual \times QE + \beta_{10} Rqual \times QL + \beta_{11} Rwcp \times QE + \beta_{12} Rwcp \times QL + u,$$

which is identical to the model presented in Section 7, except for the variables $SRqual$ and $SRwcp$ now not being included. The rational behavior can be formulated as a null hypothesis, including six parameter restrictions in the model specified above. We get the null hypothesis:

$$H_0\text{: } \beta_2 = \beta_3 = \beta_9 = \beta_{10} = \beta_{11} = \beta_{12} = 0$$

to be tested against

H_A: At least one of β_2, β_3, β_9, β_{10}, β_{11} and β_{12} is not equal to zero,

where the alternative hypothesis corresponds to a behavior of choice where the expected achieved world cup points, conditioning on the capacity of the athlete, differ for at least one of the three types of quarterfinals. A Wald test is used, which follows approximately a chi-squared distribution with 6 degrees of freedom under H_0.

It is of importance to note that these restrictions are to be tested within a reduced model, where the degree of competition, represented by the variables $SRqual$ and $SRwcp$, is left out from the model introduced in Section 7.1. In fact, for the rational behavior to exist, the degree of competition must vary among type of quarterfinals, and therefore should not be held constant by including variables capturing competition.

8.2 Testing Results for Rational Choice

Consider the estimation results from the model set out in Section 8.1.

$$\widehat{WCP} = \underset{(2.87)}{66.4^{***}} + \underset{(2.97)}{5.95^{**}} QE - \underset{(2.96)}{5.83^{**}} QL - \underset{(0.232)}{2.58^{***}} Rqual - \underset{(0.244)}{2.32^{***}} Rwcp + \underset{(0.00656)}{0.0285^{***}} Rqual \times Rwcp$$

$$+ \underset{(0.00687)}{0.0359^{***}} Rqualsq + \underset{(0.00702)}{0.0389^{***}} Rwcpsq - \underset{(0.126)}{0.238^{**}} Rqual \times QE + \underset{(0.125)}{0.156} Rqual \times QL$$

$$- \underset{(0.146)}{0.0643} Rwcp \times QE + \underset{(0.135)}{0.194^{*}} Rwcp \times QL.$$

For a top athlete, even when not conditioning on competition, the effect of going in an early quarterfinal compared to quarterfinal 3 is positive ($\hat{\beta}_2 = 5.95$), although the estimated effect has diminished somewhat. Likewise, the negative effect of a late quarterfinal is still present ($\hat{\beta}_3 = -5.83$), albeit slightly reduced. Moreover, these two effects seem to diminish in magnitude with decreased capacity ($\hat{\beta}_9 = -0.238$ and $\hat{\beta}_{11} = -0.0643$ being negative, while $\hat{\beta}_{10} = 0.156$ and $\hat{\beta}_{12} = 0.194$ being positive).

The fact that the estimated quarterfinal-type effects are smaller, when not conditioning on competition, is explained by the fact that athletes with high capacity are to some extent overrepresented in early quarterfinals and underrepresented in late quarterfinals. However, we can reject the null hypothesis of rational choice ($p < 0.001$), formulated as a certain number of restrictions of the parameters in the model, i.e., $\beta_2 = \beta_3 = \beta_9 = \beta_{10} = \beta_{11} = \beta_{12} = 0$. Hence, it could be argued that high-capacity athletes to an even larger extent should choose early quarterfinals, for the positive recovery effect to be outweighed by the negative effect of competition.

9 Individual Effects of Choice

Section 8 presented support for irrational behavior of choice at a group level. The next logical step is to examine individual effects of choice. We would like to investigate whether some athletes systematically benefit from their way of choosing quarterfinals, while others might be disfavored. The first part in this section presents the methodological framework for the analysis, while the second part shows the empirical results.

9.1 Methodology

A method to point and interval estimate the effect of choice at an individual level is proposed. The method takes as the starting point the model in Section 7. From the estimation results a typical successful choice for a high-capacity athlete would be an early quarterfinal where the competition is low, while an example of a poor choice could be a late quarterfinal with tough competition. For each athlete we aim to point estimate the effect of choice of quarterfinals on the expected achieved world cup points. Two inference approaches will be considered. Both approaches consider the uncertainty of the estimation of the parameters in the regression model set out in Section 7.1. In the first approach, yet one more source of uncertainty is modelled in that the individual's observed choice of type of quarterfinal is an outcome of a random variable. The second approach condition on the observed choices made during the competitions the individual has been participating in.

9.1.1 Approach 1

The variables in the regression model outlined in Section 7.1 are considered as random variables at an individual level. We denote for the i:th individual, $i = 1, \dots, m$, these random variables by $WCP^{(i)}$ $QE^{(i)}$, $QL^{(i)}$, $Rqual^{(i)}, \dots, SRwcp^{(i)}$. With respect to the unconditional distributions, we denote the expected values by $\mu_{WCP^{(i)}}, \mu_{QE^{(i)}}, \mu_{QL^{(i)}}, \mu_{Rqual^{(i)}}, \dots, \mu_{Rqual\times QL^{(i)}}$. For the two Bernoulli variables $QE^{(i)}$ and $QL^{(i)}$ we have $\mu_{QE^{(i)}} = p_{QE^{(i)}}$ and $\mu_{QL^{(i)}} = p_{QL^{(i)}}$, where $p_{QE^{(i)}}$ and $p_{QL^{(i)}}$ are the individual's unconditional probabilities to choose an early and a late quarterfinal, respectively. Thus, using the model specification in Section 7.1, the unconditional expected world cup points for the i:th individual can be written as

$$\mu_{WCP^{(i)}} = \beta_1 + \beta_2 p_{QE^{(i)}} + \beta_3 p_{QL^{(i)}} + \beta_4 \mu_{Rqual^{(i)}} + \beta_5 \mu_{Rwcp^{(i)}} + \beta_6 \mu_{Rqual\times Rwcp^{(i)}}$$

$$+\beta_7 \mu_{Rqualsq^{(i)}} + \beta_8 \mu_{Rwcpsq^{(i)}} + \beta_9 \mu_{Rqual\times QE^{(i)}} + \beta_{10} \mu_{Rqual\times QL^{(i)}}$$

$$+\beta_{11} \mu_{Rwcp\times QE^{(i)}} + \beta_{12} \mu_{Rwcp\times QL^{(i)}} + \beta_{13} \mu_{SRqual^{(i)}} + \beta_{14} \mu_{SRwcp^{(i)}}.$$

Eight explanatory variables are related to the choice of quarterfinal, the two dummy variables QE and QL, the four interaction variables containing QE and QL, and finally, the two variables measuring competition, $SRqual$ and $SRwcp$. Considering these eight variables, we define for the i:th individual, $i = 1, \dots, m$,

$$\theta_i = \beta_2 p_{QE^{(i)}} + \beta_3 p_{QL^{(i)}} + \beta_9 \mu_{Rqual\times QE^{(i)}} + \beta_{10} \mu_{Rqual\times QL^{(i)}}$$

$$+\beta_{11} \mu_{Rwcp\times QE^{(i)}} + \beta_{12} \mu_{Rwcp\times QL^{(i)}} + \beta_{13} \mu_{SRqual^{(i)}} + \beta_{14} \mu_{SRwcp^{(i)}}.$$

Thus, θ_i is part of $\mu_{WCP^{(i)}}$, and we interpret θ_i as the individual's unconditional expected world cup points, that is attributed to the process of choosing between quarterfinals.

Now, not conditioning on a particular individual, the unconditional expected world cup points is written as

$$\mu_{WCP} = \beta_1 + \beta_2 p_{QE} + \beta_3 p_{QL} + \beta_4 \mu_{Rqual} + \beta_5 \mu_{Rwcp} + \beta_6 \mu_{Rqual\times Rwcp}$$

$$+\beta_7 \mu_{Rqualsq} + \beta_8 \mu_{Rwcpsq} + \beta_9 \mu_{Rqual\times QE} + \beta_{10} \mu_{Rqual\times QL}$$

$$+\beta_{11} \mu_{Rwcp\times QE} + \beta_{12} \mu_{Rwcp\times QL} + \beta_{13} \mu_{SRqual} + \beta_{14} \mu_{SRwcp}.$$

Here, the parameters μ_{WCP}, p_{QE}, p_{QL}, μ_{Rqual}, …, μ_{SRwcp}, are expectations with respect to mixture distributions, all with $m = 297$ mixture components (there are 297 unique athletes making at least one choice in our data set) and mixture weights n_i/n, n_i being the number of times the i:th individual has made a choice among all $n = 2040$ choices being made during the 68 skiing sprint competitions included in the study. Thus, the parameters μ_{WCP}, p_{QE}, …, μ_{SRwcp} are defined as weighted averages of their counterparts at the individual level, i.e., $\mu_{WCP^{(i)}}, \mu_{QE^{(i)}}, \mu_{QL^{(i)}}, \mu_{Rqual^{(i)}}, \dots, \mu_{Rqual\times QL^{(i)}}$. In a similar way as θ_i was defined, we let

$$\theta = \beta_2 p_{QE} + \beta_3 p_{QL} + \beta_9 \mu_{Rqual\times QE} + \beta_{10} \mu_{Rqual\times QL}$$

$$+\beta_{11} \mu_{Rwcp\times QE} + \beta_{12} \mu_{Rwcp\times QE} + \beta_{13} \mu_{SRqual} + \beta_{14} \mu_{SRwcp}.$$

We interpret θ as the part of μ_{WCP},, which is attributed to the variables in the model that is linked to the choice of quarterfinal.

Now, define $\delta_i = \theta_i - \theta$. This difference is interpreted as the expected addition of achieved world cup points, resulting from a comparison of the i:th individual's process of choosing quarterfinals to an average good process. Thus, a positive value of δ_i means a better than average process of choosing quarterfinals, while a negative value means the opposite.

Due to certain restrictions, four out of eight parameters are known. It can be shown that,

$$p_{QE} = p_{QL} = 0.4 \text{ and } \mu_{SRqual} = \mu_{SRwcp} = 0. \quad \text{(See Appendix 2 for proof)}$$

Substitution and rearranging terms yields

$$\delta_i = \beta_2(p_{QE^{(i)}} - 0.4) + \beta_3(p_{QL^{(i)}} - 0.4) + \beta_9(\mu_{Rqual\times QE^{(i)}} - \mu_{Rqual\times QE}) +$$
$$\beta_{10}\left(\mu_{Rqual\times QL^{(i)}} - \mu_{Rqual\times QL}\right) + \beta_{11}\left(\mu_{Rwcp\times QE^{(i)}} - \mu_{Rwcp\times QE}\right) +$$
$$\beta_{12}\left(\mu_{Rwcp\times QL^{(i)}} - \mu_{Rwcp\times QE}\right) + \beta_{13}\mu_{SRqual^{(i)}} + \beta_{14}\mu_{SRwcp^{(i)}}.$$

To be more specific, a positive value of δ_i might result from individual i systematically ends up in early quarterfinals with low competition, i.e., $p_{QE^{(i)}} > 0.4$, $p_{QL^{(i)}} < 0.4$, and $\mu_{SRqual^{(i)}}$ and $\mu_{SRwcp^{(i)}}$ both being positive. (Recall that the estimates of β_2, β_{13}, and β_{14} are all positive, while the estimate of β_3 is negative.) Likewise, a negative value of δ_i could be the result of individual i often tends to compete in late quarterfinals with high competition.

Estimation of δ_i is straightforward. $\hat{\delta}_i$, the estimator of δ_i, is defined by replacing the 20 parameters by estimators in the right-hand side of the formula above, in the following way. The parameter vector $\boldsymbol{\beta}^{*\prime} = (\beta_2, \beta_3, \beta_9, \dots, \beta_{14})$ of dimension (1×8) is estimated by the vector of ordinary least squares estimators $\widehat{\boldsymbol{\beta}}^{*\prime} = (\hat{\beta}_2, \hat{\beta}_3, \hat{\beta}_9, \dots, \hat{\beta}_{14})$ used in the estimation of the regression model set out in section 7.1. The eight parameters related to a specific individual i, contained in the (1×8) dimensional vector $\boldsymbol{\gamma}_i' = \left(p_{QE^{(i)}}, \dots, \mu_{SRwcp^{(i)}}\right)$, are estimated by their sample counterparts, given by the vector $\widehat{\boldsymbol{\gamma}}_i' = \left(\hat{p}_{QE^{(i)}}, \dots, \hat{\mu}_{SRwcp^{(i)}}\right)$. For example, to estimate $p_{QE^{(i)}}$, the estimator

$$\hat{p}_{QE^{(i)}} = \frac{\sum_{j=1}^{n_i} QE_j^{(i)}}{n_i}$$

is used, where $QE_1^{(i)}, \dots, QE_{n_i}^{(i)}$ is assumed to be a random sample from a $Bern\left(p_{QE^{(i)}}\right)$, $i = 1, \dots, 297$. Likewise, the sample counterparts are used to estimate the parameter vector $\boldsymbol{\gamma}' = \left(\mu_{Rqual\times QE}, \dots, \mu_{Rwcp\times QL}\right)$, containing the four parameters from the mixture distribution, defined by $\widehat{\boldsymbol{\gamma}}' = \left(\hat{\mu}_{Rqual\times QE}, \dots, \hat{\mu}_{Rwcp\times Ql}\right)$. For example, to estimate $\mu_{Rankqual\times QE}$, we use the estimator

$$\hat{\mu}_{Rqual\times QE} = \frac{\sum_{j=1}^{n} Rqual \times QE_j}{n},$$

where $Rqual \times QE_1, \dots, Rqual \times QE_n$ is assumed to be a random sample from the mixture distribution of $Rqual \times QE$ with expectation $\mu_{Rankqual\times QE}$.

Since $\hat{\delta}_i$ is a nonlinear function of $(\widehat{\boldsymbol{\beta}}^{*\prime}, \widehat{\boldsymbol{\gamma}}', \widehat{\boldsymbol{\gamma}}_i')$ we use an approximation of the variance of $\hat{\delta}_i$, denoted by $V(\hat{\delta}_i)$. An approximation is obtained by taking the variance of a first order Taylor series expansion of this nonlinear function around the expectations $(\widehat{\boldsymbol{\beta}}^{*\prime}, \widehat{\boldsymbol{\gamma}}', \widehat{\boldsymbol{\gamma}}_i')$. For notational convenience, let $\left(\widehat{\boldsymbol{\beta}}^{*\prime}, \widehat{\boldsymbol{\gamma}}', \widehat{\boldsymbol{\gamma}}_i'\right) = \boldsymbol{\alpha}^{(i)\prime}$. We get

$$V\left(\hat{\delta}_i\right) \approx \frac{\partial \delta_i}{\partial \boldsymbol{\alpha}^{(i)\prime}} \boldsymbol{\Omega}^{(i)} \frac{\partial \delta_i}{\partial \boldsymbol{\alpha}^{(i)}},$$

where $\boldsymbol{\Omega}^{(i)} = V\left(\widehat{\boldsymbol{\alpha}}^{(i)}\right)$ is the variance-covariance matrix. We assume $\boldsymbol{\Omega}^{(i)}$ to be block diagonal,

$$\boldsymbol{\Omega}^{(i)} = \begin{bmatrix} \boldsymbol{\Omega}_{11} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Omega}_{22} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \boldsymbol{\Omega}_{33}^{(i)} \end{bmatrix},$$

where the matrix $\boldsymbol{\Omega}_{11} = V(\widehat{\boldsymbol{\beta}}^*)$ has dimension (8×8), derived under the assumption of general heteroskedasticity of the error term u in the regression model from Section 7.1 (see White (1980)), $\boldsymbol{\Omega}_{22} = V(\widehat{\boldsymbol{\gamma}})$, and $\boldsymbol{\Omega}_{33}^{(i)} = V(\widehat{\boldsymbol{\gamma}}_i)$, the latter two with dimensions (4×4) and (8×8), respectively. The assumption of the off-diagonal matrices $\boldsymbol{\Omega}_{12}$, $\boldsymbol{\Omega}_{21}$, $\boldsymbol{\Omega}_{13}^{(i)}$, and $\boldsymbol{\Omega}_{31}^{(i)}$ all being $\mathbf{0}$ is not a controversial assumption. There is undoubtedly a dependency between $\widehat{\boldsymbol{\gamma}}$ and $\widehat{\boldsymbol{\gamma}}_i$, since the observations from the

i:th individual is included in the calculations of both $\hat{\boldsymbol{\gamma}}$ and $\hat{\boldsymbol{\gamma}}_i$. Therefore, the assumption of $\boldsymbol{\Omega}_{23}^{(i)}$ and $\boldsymbol{\Omega}_{32}^{(i)}$ being $\mathbf{0}$ might be questioned. Yet, the number of observations for the i:th individual constitutes a small proportion of the total number of observations used for $\hat{\boldsymbol{\gamma}}$, meaning that the dependency is small, and therefore is neglected ($\max\{n_i\} = 32$ compared to $n = 2040$).

Appendix 3 contains full information on the elements of the matrix $\boldsymbol{\Omega}^{(i)}$. The elements of the vector $\frac{\partial \delta_i}{\partial \boldsymbol{\alpha}^{(i)\prime}}$ is given, as well.

By applying the Delta method, the distribution of $\hat{\delta}_i$ can be approximated. We get, for sufficiently large sample sizes of n_i and n,

$$\hat{\delta}_i \overset{appr}{\sim} N\left(\delta_i, \frac{\partial \delta_i}{\partial \boldsymbol{\alpha}^{(i)\prime}} \boldsymbol{\Omega}^{(i)} \frac{\partial \delta_i}{\partial \boldsymbol{\alpha}^{(i)}}\right).$$

The result follows from the elements in the vector of estimators $\hat{\boldsymbol{\alpha}}^{(i)\prime}$, all being unbiased and asymptotically normal.

An approximate $100(1-\alpha)$ % confidence interval estimator for δ_i is now given by

$$\left(\hat{\delta}_i - z_{\frac{\alpha}{2}} \sqrt{\widehat{\frac{\partial \delta_i}{\partial \boldsymbol{\alpha}^{(i)\prime}}} \widehat{\boldsymbol{\Omega}}^{(i)} \widehat{\frac{\partial \delta_i}{\partial \boldsymbol{\alpha}^{(i)}}}}\,;\ \hat{\delta}_i + z_{\frac{\alpha}{2}} \sqrt{\widehat{\frac{\partial \delta_i}{\partial \boldsymbol{\alpha}^{(i)\prime}}} \widehat{\boldsymbol{\Omega}}^{(i)} \widehat{\frac{\partial \delta_i}{\partial \boldsymbol{\alpha}^{(i)}}}}\right),$$

where $z_{\alpha/2}$ is the $100(1-\alpha/2)$th percentile for the standard normal distribution, while $\widehat{\boldsymbol{\Omega}}^{(i)}$ and $\widehat{\frac{\partial \delta_i}{\partial \boldsymbol{\alpha}^{(i)\prime}}}$ are estimators for $\boldsymbol{\Omega}^{(i)}$ and $\frac{\partial \delta_i}{\partial \boldsymbol{\alpha}^{(i)\prime}}$, respectively. To get $\widehat{\boldsymbol{\Omega}}^{(i)}$, the elements in $\boldsymbol{\Omega}^{(i)}$ are replaced by corresponding sample moments, while $\widehat{\frac{\partial \delta_i}{\partial \boldsymbol{\alpha}^{(i)\prime}}}$ is obtained by replacing $(\boldsymbol{\beta}^{*\prime}, \boldsymbol{\gamma}', \boldsymbol{\gamma}_i')$ by $(\hat{\boldsymbol{\beta}}^{*\prime}, \hat{\boldsymbol{\gamma}}', \hat{\boldsymbol{\gamma}}_i')$ in the elements of $\frac{\partial \delta_i}{\partial \boldsymbol{\alpha}^{(i)\prime}}$.

9.1.2 Approach 2

In the second approach we condition on the observed choices made during the competitions the individual has been participating in. It means that the analysis is conditioned on the estimates $\hat{\boldsymbol{\gamma}}$ and $\hat{\boldsymbol{\gamma}}_i$. Therefore, in the expression for δ_i the actual estimates $\hat{\boldsymbol{\gamma}}$ and $\hat{\boldsymbol{\gamma}}_i$ replace the parameters $\boldsymbol{\gamma}$ and $\boldsymbol{\gamma}_i'$ and the interpretation is somewhat different. The parameter, to be denoted by δ_i^*, is interpreted as the expected addition of achieved world cup points per competition, as a result of the actual choice – not the process – of the i:th individual during the study period compared to the observed average good choice for the same period. Since the only source of uncertainty of the estimate $\hat{\delta}_i^*$ comes from estimation of $\boldsymbol{\beta}^*$, $V(\hat{\delta}_i^*) < V(\hat{\delta}_i)$, and we get

$$\hat{\delta}_i \overset{appr}{\sim} N\left(\delta_i, \frac{\partial \delta_i}{\partial \boldsymbol{\beta}'} \boldsymbol{\Omega}_{\mathbf{11}} \frac{\partial \delta_i}{\partial \boldsymbol{\beta}}\right),$$

and an approximate $100(1-\alpha)$ % confidence interval estimator for δ_i is

$$\left(\hat{\delta}_i - z_{\alpha/2} \sqrt{\widehat{\frac{\partial \delta_i}{\partial \boldsymbol{\beta}'}} \widehat{\boldsymbol{\Omega}}_{\mathbf{11}} \widehat{\frac{\partial \delta_i}{\partial \boldsymbol{\beta}}}}\,;\ \hat{\delta}_i + z_{\alpha/2} \sqrt{\widehat{\frac{\partial \delta_i}{\partial \boldsymbol{\beta}'}} \widehat{\boldsymbol{\Omega}}_{\mathbf{11}} \widehat{\frac{\partial \delta_i}{\partial \boldsymbol{\beta}}}}\right).$$

Expressions for the estimators $\widehat{\boldsymbol{\Omega}}_{\mathbf{11}}$ and $\widehat{\frac{\partial \delta_i}{\partial \boldsymbol{\beta}'}}$ are obtained analogously to $\widehat{\boldsymbol{\Omega}}^{(i)}$ and $\widehat{\frac{\partial \delta_i}{\partial \boldsymbol{\alpha}^{(i)\prime}}}$, respectively, used in the first approach.

9.2 Results of Individual Effects on Choice

In this section we present point and interval estimates of individual effects of choice based on both approaches described in section 9.1.

9.2.1 Approach 1

Consider Figure 3. It shows point and interval estimates for confidence intervals of δ_i for those *115* athletes having participated in more than five competitions. For only nine athletes we can be at least 95 percent confident that their way of choosing quarterfinals in the long run is better than average (three athletes) or worse than average (six athletes). In case $\delta_i = 0$ for all 115 athletes, the expected number of intervals not containing zero is 5.75 and from the binomial distribution we get a probability of 0.12 to get nine intervals or more not covering zero in that case. Thus, there is some support for at least one athlete to have a parameter value $\delta_i \neq 0$, although the support is not strong.

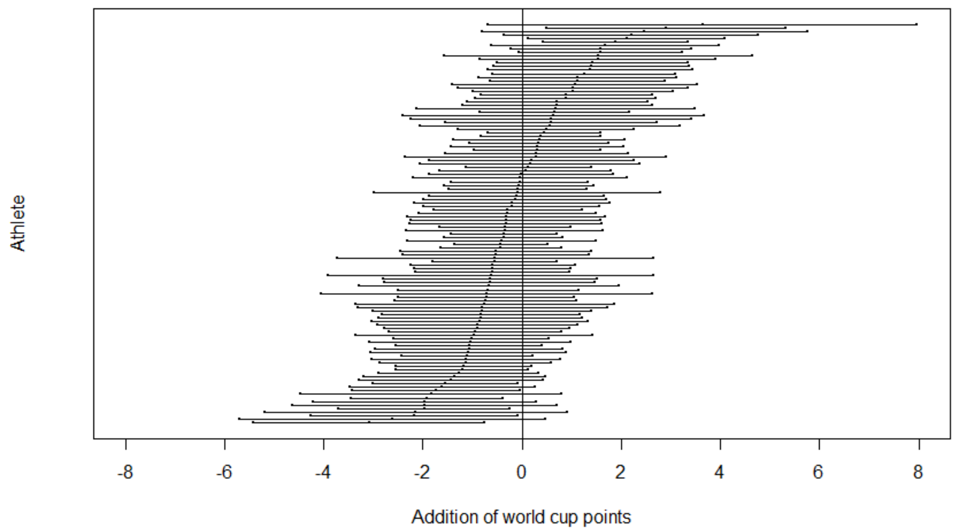

Figure 3. 95 percent confidence intervals for δ_i for those 115 athletes having participated in more than five competitions. Point estimates dotted in the middle of each interval.

The wide confidence intervals are, foremost, a result of two sources of uncertainty. First, there is uncertainty in the estimation of the beta parameters, especially the parameters corresponding to the recovery effect, where the standard errors are quite high. Second, since many athletes have participated in few competitions, the distribution of their individual choice of type of quarterfinal is also estimated with a lot of uncertainty.

9.2.2 Approach 2

In Figure 4 we show point and interval estimates of δ_i^* for the same 115 athletes. The confidence intervals are not as wide as the previous ones since we now condition on the choice and the only source of uncertainty comes from the estimation of the beta parameters.

Here, 34 athletes have intervals not covering zero. For twelve out of these athletes, we are at least 95 percent confident that their actual choice of type of quarterfinal has had an expected positive effect on their performance in terms of achieved world cup points. For 22 athletes the effect is expected to be negative.

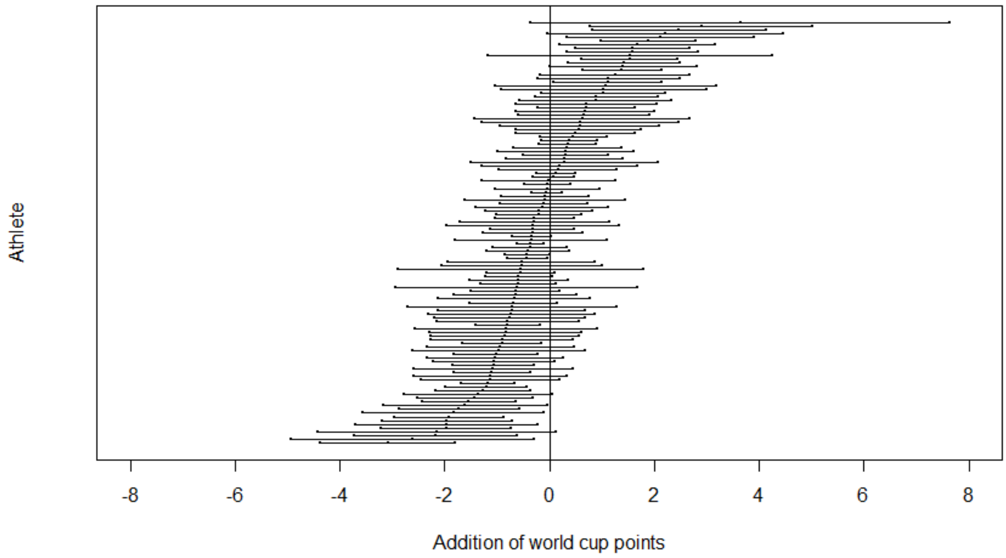

Figure 4: 95 percent confidence intervals for δ_i^* for those 115 athletes having participated in more than 5 competitions. Point estimates dotted in the middle of each interval.

Thus, we can identify far more losers than winners within this design of assigning quarterfinals. How come we have this unbalance, is this not a zero-sum game, where we expect the number of athletes benefiting from the design to approximately equal the number of athletes losing from it? Recall from the results in section 7.2 that high-capacity athletes should to an even larger extent choose early quarterfinals, for the two effects of competition and recovery to cancel each other out. Thus, high-capacity athletes are most likely overrepresented among those athletes losing from the design. In addition, it is easier to identify high-capacity athletes in this process, since the expected effect of their choice of quarterfinal is larger than for low-capacity athletes. These two things taken together explain the unbalanced result.

10 Conclusions and Discussion

The first part of our work is related to Karlsson and Lunander (2020), the main difference being that we adopt another objective function when predicting and empirically estimating the effect of athletes' choices of races in a knock-out tournament. When deciding which of the five quarterfinals to compete in, the athlete in our analysis is not assumed to maximize her probability to reach the podium, but instead is seeking to maximize her expected achieved world cup points. We believe that for most of the 30 qualified athletes choosing their quarterfinals, this objective is more appropriate than primarily chasing the podium. Nevertheless, our obtained results are in line with those presented in Karlsson and Lunander (2022). We find that a high ranked athlete is predicted to choose an early rather than a late race, a result being consistent with data. The outcome from the regression analysis, using the athlete's achieved world cup points as the dependent variable, shows that the choice of an early quarterfinal has a positive significant effect upon the athlete's achieved world point. In addition, the expected achieved world cup points are still found to be higher when choosing an early quarterfinal, conditioning on athletes' capacity, despite the impact of increased competition in the early quarterfinals, suggesting sub-optimal decisions as a group. Also, the athlete's rank from the qualification round is still the strongest predictor to determine achieved world cup points

The major contribution of our work is found in its second part, where we developed a method to assess whether some athletes are more or less tactical skilled than others in managing the balance between recovery and expected competition when choosing their quarterfinals. At an individual level, for twelve out of 115 athletes, we are at least 95 percent confident that their actual choice of type of quarterfinal has had an expected positive effect on their performance in terms of achieved world cup points. For 22 athletes the effect is expected to be negative. The estimated individual effects ranges from around minus three to plus four points. However, when we don't condition on their actual choices, i.e., when we add the extra uncertainty implied by viewing the choices made as outcomes of random variables, we only have a moderate support for the hypothesis that some athletes in the long run choose quarterfinals in a way that has an impact - positive or negative - on achieved world cup points.

In our analysis, we have not got into the question on what cognitive processes are underpinning the athletes' choices of quarterfinals. A future extension of this work would be to relate our observed decision making to the concept of simple heuristics or rules of thumb, well suited in sport, where athletes in many situations face limited time and information. While there exists a huge body of literature focusing on the role of heuristics in overall decision making, the number of similar studies in sports is limited (e.g., Raab, 2012; Raab et al., 2019).

References

Kalén, A., Bisagno, E., Musculus, L., Raab, M., Pérez-Ferreirós, A., Williams, A. M., ... & Ivarsson, A. (2021). The role of domain-specific and domain-general cognitive functions and skills in sports performance: A meta-analysis. *Psychological bulletin*, *147*(12), 1290.

Karlsson, N., Lunander, A. (2020). Choosing Opponents in Skiing Sprint Elimination Tournaments. *Working paper 6:2020,* Business School at Örebro University,

Mann, D. T., Williams, A. M., Ward, P., & Janelle, C. M. (2007). Perceptual-cognitive expertise in sport: A meta-analysis. *Journal of Sport and Exercise Psychology*, *29*(4), 457-478.

Raab, M. (2012). Simple heuristics in sports. *International Review of Sport and Exercise Psychology*, *5*(2), 104-120.

Raab, M., MacMahon, C., Avugos, S., & Bar-Eli, M. (2019). Heuristics, biases, and decision making . *in* Williams, A. M., & Jackson, R. C. (Eds.). (2019). *Anticipation and decision making in sport*. Abingdon: Routledge.

Scharfen, H. E., & Memmert, D. (2019). Measurement of cognitive functions in experts and elite athletes: A meta-analytic review. *Applied Cognitive Psychology*, *33*(5), 843-860.

Voss, M. W., Kramer, A. F., Basak, C., Prakash, R. S., & Roberts, B. (2010). Are expert athletes 'expert'in the cognitive laboratory? A meta-analytic review of cognition and sport expertise. *Applied cognitive psychology*, *24*(6), 812-826.

Appendix 1

Proof of Proposition 1

Given that player B initially chooses the first semi-final (s.1), we get

$$E_{A,1} = p_{A,1}^{(1)} + kp_{A,1}^{(2)},$$

where

$$p_{A,1}^{(1)} = 0.5 \times 0.5(1-(1-p)c) + 0.5 \times 0.5(1-(1-p)c)$$

and

$$p_{A,1}^{(2)} = k(0.5 \times 0.5(1-p)c + 0.5 \times 0.5(1-p)c.$$

The first part of $p_{A,1}^{(1)}$ is the probability that player A beats player B, player C beats player D in s.2, and in the final player A beats player C. Note that player C's probability of beating player A in the final, $(1-p)c$, is reduced by the factor c. The second part identifies the same probabilities as the first part, but now it is player D who advances to the final. The first part of $p_{A,1}^{(2)}$ is the probability that player A beats player B, player C beats player D in s.2, and in the final, player A loses to C. The second part identifies the same probabilities as the first part, except for C and D are changing places.

If player A instead chooses s.2, then player C is better off choosing s.1 than s.2, since it can be shown that $E_{C,2} - E_{C,3} = (1-p)(1-c)(1-k) > 0$.

Thus, the expected achieved world cup points for A given that player B chooses s.1 and player A chooses s.2 can be written as

$$E_{A,2} = p \times p0.5c + p(1-p)pc + k[p \times p(1-0.5c) + p(1-p)(1-pc)].$$

Player A will choose to compete against player B in s.1 if $E_{A,1} - E_{A,2} > 0$. Using the expressions for $E_{A,1}$and $E_{A,2}$, we obtain the condition for player A choosing to compete against player B in s.1 as $c < \frac{0.5-pk}{0.5-0.5p+1.5p^2-p^3(1-k)}$.

Now, given this condition, we get

$$E_{B,1} = E_{A,1}$$

due to symmetry. Otherwise, player B will meet player C in s.1, resulting in

$$E_{B,2} = p \times p(1-0{,}5c) + p(1-p)(1-(1-p)c) + k(p \times p0.5c + p(1-p)(1-p)c).$$

Turning to the case where player B initially chooses the second semi-final (s.2). If player A chooses s.1, it can easily be verified that player C prefers to compete against player A in s.1 rather than facing player B in s.2. Actually, $E_{C,4} - E_{C,5} = (1-p)(1-c)(1-k) > 0$.

Thus, again for symmetric reasons, we get

$$E_{A,4} = E_{B,2}.$$

In fact, A prefers to meet C in s.1 rather than B in s.2. To prove that we make use of the two inequalities $p_{A,4}^{(1)} > p_{A,6}^{(1)}$ and $p_{A,4}^{(1)} + p_{A,4}^{(2)} > p_{A,6}^{(1)} + p_{A,6}^{(2)}$. We get, by substitution and using that $1-k > 0$,

$$E_{A,4} = p_{A,4}^{(1)} + kp_{A,4}^{(2)} > p_{A,4}^{(1)} + k\left(p_{A,6}^{(1)} + p_{A,6}^{(2)} - p_{A,4}^{(1)}\right) =$$

$$p_{A,4}^{(1)}(1-k) + kp_{A,6}^{(1)} + kp_{A,6}^{(2)} > p_{A,6}^{(1)}(1-k) + kp_{A,6}^{(1)} + kp_{A,6}^{(2)} =$$

$$p_{A,6}^{(1)} + kp_{A,6}^{(2)} = E_{A,6}.$$

To summarize, if player B initially chooses s.2 then he will face player D in this semifinal. Thus, player B's expected achieved world cup points when choosing s.2 is

$$E_{B,4} = E_{A,1}.$$

However, if all players act to maximize the expected world cup points, player B will never choose s.2. For the case $c < \frac{0.5-pk}{0.5-0.5p+1.5p^2-p^3(1-k)}$, it is relevant for player B to compare $E_{B,1}$ with $E_{B,4}$. For this case, we have earlier found that $E_{A,1} > E_{A,2}$. Since we also have $E_{A,1} = E_{B,1}$ and $E_{A,2} = E_{B,4}$, we get $E_{B,1} > E_{B,4}$.

For the case $c > \frac{0.5-pk}{0.5-0.5p+1.5p^2-p^3(1-k)}$ we compare $E_{B,2}$ with $E_{B,4}$. Using the constraints $p_{B,2}^{(1)} > p_{B,4}^{(1)}$ and $p_{B,2}^{(1)} + p_{B,2}^{(2)} = p_{B,4}^{(1)} + p_{B,4}^{(2)}$, it can be verified that $E_{B,2} > E_{B,4}$.

Appendix 2

To prove $p_{QE} = 0.4$, we note that the mixture distribution of QE is $Bern(p_{QE})$, where

$$p_{QE} = \sum_{i=1}^{297} \frac{n_i}{2040} p_{QE^{(i)}}.$$

Now, define $QE_1^{(1)}, \ldots, QE_{n_1}^{(1)}, \ldots, QE_{297}^{(297)}, \ldots, QE_{n_{297}}^{(297)}$, conditioning on $n_1, \ldots, n_{297}$, to be a random sample from the mixture distribution of QE. Since the total number of choices of early quarterfinals made is 816 (0.4×2040), we get the restriction

$$\sum_{i=1}^{297} \sum_{j=1}^{n_i} QE_j^{(i)} = 816.$$

Dividing by 2040 and taking the expectation of both sides, we get

$$\sum_{i=1}^{297} \frac{n_i}{2040} p_{QE^{(i)}} = 0.4,$$

where the left-hand side defines p_{QE}. In a similar way, $p_{QL} = 0.4$ can be proved.

To prove $\mu_{SRqual} = 0$, we note that the expectation of the mixture distribution of $SRqual$ is given by

$$\mu_{SRqual} = \sum_{i=1}^{297} \frac{n_i}{2040} \mu_{SRqual^{(i)}}$$

Now, define $SRqual_1^{(1)}, \ldots, SRqual_{n_1}^{(1)}, \ldots, SRqual_1^{(297)}, \ldots, SRqual_{n_{297}}^{(297)}$, conditioning on $n_1, \ldots, n_{297}$, to be a random sample from the mixture distribution of $SRqual$. Considering the way in which the variable is defined, as a deviation from 93, the average sum of ranking numbers, we get the restriction

$$\sum_{i=1}^{297} \sum_{j=1}^{n_i} SRqual_j^{(i)} = 0.$$

Dividing by 2040 and taking the expectation of both sides, we get

$$\sum_{i=1}^{297} \frac{n_i}{2040} \mu_{SRqual^{(i)}} = 0.4,$$

where the left-hand side defines μ_{SRqual}. In a similar way it can be proved that $\mu_{SRwcp} = 0$.

Appendix 3

Below the elements of $\boldsymbol{\Omega}_{22}$ and $\boldsymbol{\Omega}_{33}^{(i)}$ are given. For the elements of $\boldsymbol{\Omega}_{11}$, the reader is referred to White (1980).

The (4×4) matrix $\boldsymbol{\Omega}_{22}$ is given by

$$\boldsymbol{\Omega}_{22} = \begin{bmatrix} \frac{V(Rqual\times QE)}{n} & \frac{C(Rqual\times QE, Rqual\times QL)}{n} & \cdots & \frac{C(Rqual\times QE, Rwcp\times QL)}{n} \\ \frac{C(Rqual\times QE, Rqual\times QL)}{n} & \frac{V(Rqual\times QL)}{n} & \cdots & \frac{C(Rqual\times QL, wcp\times QL)}{n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{C(Rqual\times QE, Rwcp\times QL)}{n} & \frac{C(Rqual\times QL, wcp\times QL)}{n} & \cdots & \frac{V(Rwcp\times QL)}{n} \end{bmatrix},$$

while the (8×8) matrix $\boldsymbol{\Omega}_{33}^{(i)}$ is defined by

$$\boldsymbol{\Omega}_{33}^{(i)} = \begin{bmatrix} \frac{V(QE^{(i)})}{n_i} & \frac{C(QE^{(i)}, QL^{(i)})}{n_i} & \cdots & \frac{C(QE^{(i)}, SRwcp^{(i)})}{n_i} \\ \frac{C(QE^{(i)}, QL^{(i)})}{n_i} & \frac{V(QL^{(i)})}{n_i} & \cdots & \frac{C(QL^{(i)}, SRwcp^{(i)})}{n_i} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{C(QE^{(i)}, SRwcp^{(i)})}{n_i} & \frac{C(QL^{(i)}, SRwcp^{(i)})}{n_i} & \cdots & \frac{V(SRwcp^{(i)})}{n_i} \end{bmatrix}.$$

The elements of the vector of partial derivatives $\frac{\partial \delta_i}{\partial \boldsymbol{\alpha}^{(i)\prime}}$ are given below. First, the derivatives with respect to the parameter vector $\boldsymbol{\beta}^{*\prime}$ are

$$\frac{\partial \delta_i}{\partial \beta_2} = p_{QE^{(i)}} - 0.4; \frac{\partial \delta_i}{\partial \beta_3} = p_{QL^{(i)}} - 0.4; \frac{\partial \delta_i}{\partial \beta_9} = \mu_{Rqual\times QE^{(i)}} - \mu_{Rqual\times QE};$$

$$\frac{\partial \delta_i}{\partial \beta_{10}} = \mu_{Rqual\times QL^{(i)}} - \mu_{Rqual\times QL}; \frac{\partial \delta_i}{\partial \beta_{11}} = \mu_{Rwcp\times QE^{(i)}} - \mu_{Rwcp\times QE};$$

$$\frac{\partial \delta_i}{\partial \beta_{12}} = \mu_{Rwcp\times QL^{(i)}} - \mu_{Rwcp\times QL}; \frac{\partial \delta_i}{\partial \beta_{13}} = \mu_{SRqual^{(i)}}; \frac{\partial \delta_i}{\partial \beta_{14}} = \mu_{SRwcp^{(i)}}.$$

Second, for the parameter vector $\boldsymbol{\gamma}'$ we get

$$\frac{\partial \delta_i}{\partial \mu_{Rqual\times QE}} = -\beta_9; \frac{\partial \delta_i}{\partial \mu_{Rqual\times QL}} = -\beta_{10}; \frac{\partial \delta_i}{\partial \mu_{Rwcp\times QE}} = -\beta_{11}; \frac{\partial \delta_i}{\partial \mu_{Rwcp\times QL}} = -\beta_{12}.$$

Third, the partial derivatives with respect to $\boldsymbol{\gamma}_i'$ are found to be

$$\frac{\partial \delta_i}{\partial p_{QE^{(i)}}} = \beta_2; \frac{\partial \delta_i}{\partial p_{QL^{(i)}}} = \beta_3; \frac{\partial \delta_i}{\partial \mu_{Rqual\times QE^{(i)}}} = \beta_9; \frac{\partial \delta_i}{\partial \mu_{Rqual\times QL^{(i)}}} = \beta_{10}; \frac{\partial \delta_i}{\partial \mu_{Rwcp\times QE^{(i)}}} = \beta_{11}; \frac{\partial \delta_i}{\partial \mu_{Rwcp\times QL^{(i)}}} = \beta_{12};$$

$$\frac{\partial \delta_i}{\partial \mu_{SRqual^{(i)}}} = \beta_{13}; \frac{\partial \delta_i}{\partial \mu_{SRwcp^{(i)}}} = \beta_{14}.$$