ERASMUS SCHOOL OF ECONOMICS

Dispersion within sports teams

An empirical study of sports data from the National Beach Volleyball Competition in the Netherlands.

Nathalie Boerefijn (347301)

29-7-2016

This thesis was the final step in the completion of my Master in Entrepreneurship and Strategy at the School of Economics, Erasmus University Rotterdam. This thesis would not exist without the guidance of my thesis coach, Thomas Peeters. Many thanks to him for his support and patience (especially when the process slowed down enormously since I got a job).
Special thanks to my parents who enabled me to obtain my MSc degree in the first place. Specifically, my dad, who taught me how to scrape text from websites with a programming technique. Without this, it would be impossible to obtain the datasets.
Finally, I would like to thank my brothers, roommates, friends and colleagues for their loving support. Their belief in my ability kept me going and gave me the encouragement to successfully complete my thesis and obtain my master degree.

ABSTRACT

Data from the National Beach Volleyball Competition of the Netherlands is used to investigate the effect of dispersion of ability, age and length within beach volleyball teams. Dispersion of players' abilities leads to better outcomes in beach volleyball tournaments. This indicates that low-ability players gain and learn from high-ability players. Dispersion of players' age, however, does not improve the outcome of a team in a tournament. This paper found that the outcome raises if two older (and therefore more experienced) players form a team. With regard to dispersion in length, no change in outcome is found. A minimum length is required to play professional beach volleyball, and the variation between all these tall players do not show any effect in the results off a tournament.

1 INTRODUCTION

"Team organization is pervasive in the workplace" (thus Bandiera, Barankay & Rasul, 2013, p. 1079). Teamwork could improve worker's productivity and is therefore important within organizations (Bandiera, Barankay & Rasul, 2013). Earlier, Nickerson & Owan (2003) also found that the adoption of teams in organizations improves worker's productivity. These studies raise the question under which conditions teams improve worker's productivity in general. To answer this question, many empirical studies have examined to what extent the productivity of workers depends on the team composition.

Team composition is driven by the two components workers' abilities and social connections, if team members are free to choose their own team members (Bandiera et al, 2013). However, introducing team incentives drives high-ability workers into forming teams with each other, while social connections do not seem to be relevant anymore. Therefore, workers' ability is a decisive factor in team composition. Nickerson & Owan (2003) contributed to the team composition theory by comparing the ability-levels of workers within a team. They found that teams with more heterogeneous abilities are more productive.

Economic literature describes this heterogeneity of characteristics, such as workers' ability and skill levels, within a team as *the dispersion of workers*. In 2008 Iranzo, Schivardi, & Tosetti contributed to the team composition theory by studying dispersion of the worker's skill level among team members. They found that equal workers, which are workers in the same group (production or non-production workers), maximize productivity if their skill levels are dispersed. Both dispersion of workers' ability and workers' skill levels could therefore enhance the team's overall productivity.

This study contributes to the team composition theory by examining the effects of dispersion of the abilities among two team members on the performance of the team. In addition to the dispersion of the players' ability, the dispersion of some other characteristics of the two players will be examined as well, i. e. age and length of the team members. This study is unique, because it uses quantitative sports data from the professional National Beach Volleyball Competition of the Netherlands. Since the outcomes of beach volleyball tournaments are an indication for the performance of a beach volleyball player, it creates a similar measure as the worker's productivity of an employee within an organization.

The use of sports data to examine economic theories provides several advantages. First of all, the activities and performance of a player, as well as the team performance is observable by everyone (Borland & Lye, 1996). Besides, Borland & Lye (1996) mentioned the possibility of collecting data from a reasonable lengthy period of time. D'Addona & Kind (2012) stated that performances based on match outcomes in the sports market are simpler, more reliable and less controversial. They agree with Borland & Lye (1996) that the performance in sports markets is far more visible and mention that outcomes of sports games cannot be manipulated easily. Moreover, d'Addona & Kind (2012) argue that data in the sports market is available in a more frequent way. Matches within sports are mostly played on a weekly basis, which is far more frequent than most quarterly results in businesses. Both Yang et al. (2009) and d´Adonna & Kind (2012) agree on the advantages of using the sports market mentioned by Borland & Lye (1996). Additionally, they mention that the relatively high secrecy within economic markets is considered to be a major problem and it would therefore be easier to construct a dataset for a more robust empirical study for the sports market instead.

The National Beach Volleyball Competition of the Netherlands is comparable to the incentive scheme in the form of a tournament that Bandiera et al. (2013) analysed in their field experiment. The first similarity is the fact that both the workers in the experiment and the beach volleyball players are free to choose their team members. Secondly, the National Beach Volleyball Competition, as well as the last part of the experiment are a tournament, where the ‘winners’ receive a monetary prize. Lastly, conforming to the findings in the experiment by Bandiera et al. beach volleyball players are in general forced to choose a partner with high-ability to be able to enter the National Beach Volleyball Competition. Which means that an incentive scheme in the form of a tournament generally eliminates team composition based on social connections. One of the differences between the experiment and the National Beach Volleyball Competition is the fact that the teams in the experiment contain five workers and beach volleyball teams only consist of two players. (Bandiera et al., 2013; NeVoBo, 2013)

The aim of this empirical study is to find the team composition in beach volleyball that maximizes the outcome, i.e. the position of the team’s ranking, in a tournament, answering the following research question:

What is the effect dispersion of the players’ ability, age and length on the outcome in the professional National Beach Volleyball Competition in the Netherlands?

In addition to the work of Nickerson & Owan (2003) and Iranzo et al. (2008), this study expects that dispersed abilities, ages and lengths of the two beach volleyball players that form a team will have a positive effect on the outcome of the tournament. This will be tested by using a quantitative dataset, which contains one-to-one two-sided matched players (as described by Fox in 2006) and their characteristics as well as their results in the National Beach Volleyball Competition. The following hypotheses are formed:

> *H1:* dispersion of the players' ability will have a positive effect on the outcome in a tournament.
> *H2:* dispersion of the age of the players will have a positive effect on the outcome in a tournament.
> *H3:* dispersion of the length of the players will have a positive effect on the outcome in a tournament.

Chapter two describes the theoretical framework, in which the dispersion theory will be discussed and the team composition theory will be more extensively described. The third chapter will describe the National Beach Volleyball Competition in the Netherlands. Then, the elements expected to influence the outcome of a beach volleyball team in a tournament will be illustrated. After that the chapter explains the three hypotheses. Further, chapter four describes the dataset and the methods used, while chapter five explains the results. Lastly, the sixth chapter draws the conclusions.

2 THEORETICAL FRAMEWORK

The theoretical literature widely describes how the adoption of teams in organizations can improve worker's productivity (Alchian & Demsetz, 1972; Hamilton, Nickerson & Owan, 2003; Bandiera et al., 2013). Depken (2000) defines a team as two or more workers who independently or together perform specific tasks to come to a final output. The workers who form a team can be homogeneous or heterogeneous and have dispersed characteristics.

2.1 THE EFFECT OF DISPERSION DESCRIBED IN PREVIOUS LITERATURE

Nickerson & Owan (2003) designed an experiment for the Koret company in the United States. In this company the production process changed between 1995 and 1997 from individual to team production. The researchers tested the effect of team composition on worker's productivity by focussing on the heterogeneity or homogeneity of workers within teams. Nickerson & Owan found that more heterogeneous teams are more productive. This is either caused by a higher productivity norm, because of the workers with a high-ability in the team or it is caused by mutual learning, which means that workers with a higher ability teach workers with a lesser ability to increase productivity. (Nickerson & Owan, 2003)

Most research on the heterogeneity or dispersion of workers concentrates on the dispersion of the wages of workers. Research shows that dispersion of the wages of the workers could have a negative impact on the economic outcome. Firstly, in 1991 Levine found that less dispersed wages of workers increase productivity of the workers, because of tighter cohesion among the workers. Subsequently, the effect of wage dispersion on productivity has been examined for players in the National Basketball Association (NBA), the Major League Baseball (MLB) and the Major League Soccer (MLS) of the United States. Both Berri & Jewell (2004) and Katayama & Nuch (2011) did not find any significant effects for players of the NBA. However, for players from the MLB and the MLS researchers did find significant outcomes (Coates et al, 2014). For the MLB, Bloom (1999) found a reduction in individual and group performance and Depken (2000) found a reduction in team performance for a more dispersed workforce. In addition to that, Coates et al. (2014) found a negative relation between dispersion and team performance in the MLS. However, the number of observations in this article is relatively low, which consequently leads to questionable results.

Contrary to these results, Ramaswamy and Rowthorn (1991) found that dispersion of the worker's wage should have a positive impact on the economic outcome.

In addition to the research on worker's wages dispersion, Iranzo et al. (2008) apply dispersion theory on the skill level of workers. Iranzo et al. (2008) contributed to the team composition theory by looking at the effect of a dispersed workforce on the productivity in the Italian labour market. Furthermore, Iranzo et al. (2008) distinguished between worker's skills that are complements and worker's skills that are substitutes. On the one hand, Iranzo et al. (2008) stated that combining workers with similar skills could lead to an optimal outcome, which is the case if workers are complements. On the other hand, the authors argue that it is useful to have a substitutable workforce with some very talented high-skilled workers augmented with low-skilled workers. The results of the article of Iranzo et al. (2008) says that it is optimal to match production and non-production workers with approximately the same skill level. However, they also find that workers within the production or non-production group maximize productivity if their skill levels are dispersed.

2.2 TEAM COMPOSITION DESCRIBED BY MATCHING GAMES

In 2006 Fox introduced some models that measure the relative importance of different characteristics of individuals that are matched. This model could be used in a one-to-one, one-to-many and a many-to-many matched market. In a one-to-one two-sided matching market, the situation is as follows: a definite number of individuals form self-chosen pairs (Fox, 2006), which describes exactly what happened in the experiment of Bandiera et al. (2013). During this experiment, workers of a leading soft fruit producer in the United Kingdom were free to choose their own team members every week. The team in this experiment consists of 5 team members. Therefore, the experiment is an example of a two-sided many-to-one matching market.

Matching games are extensively described in economic theory (Becker, 1973). Matching could occur in all kinds of markets where two agents choose their partners strategically (Yang, Shi & Goldfarb, 2009). Some examples of these markets are the labour market (Haskel, Hawkes, & Pereira, 2005; Hellerstein & Neumark, 2007; Iranzo et al., 2008), the marriage market (Becker, 1973; Choo & Siow, 2006; Dupuy & Galichon, 2014) and the sports market (Borland & Lye, 1996; Yang, Shi & Goldfarb, 2009).

Becker (1973) was the first one to use a two-sided matching model to empirically test the matching between men and women (Fox, 2006). Since each man and each woman competes and tries to maximize his or her utility by finding the best partner, Becker (1973) argues that market conditions exist and marriage could be seen as a market. The marriage market has a major influence on the economy through for example population growth, labour-force participation of women, inequality in income, and ability (Becker, 1973). However, the disadvantage of the marriage market is the fact that no direct economic output is generated by marriages.

A clear example of an one-to-one two-sided matching market, which generates economic output, is employer-employee matching in the labour market. Hellerstein & Neumark (2007) have used a matched employer-employee dataset from the United States and Haskel et al. (2005) for the United Kingdom. The aim of the research by Hellerstein & Neumark (2007) and Haskel et al. (2005) is to find workers' characteristics that have a positive effect on productivity.

Despite of the existence of economic output in the labour market, the problem of measuring performance in such a market has been explained in numerous articles. To overcome these problems, various researchers used the sports market to gather more reliable outputs. Borland & Lye (1996) for example, contribute to the employer-employee matching problem by using data from the Australian Football market to measure coach-team matching effects on the team's performance. Moreover, Yang et al. (2009) used a two-sided one-to-one matching model by Fox (2006) to measure how matches between professional athletes and teams maximize their total added value created by brand alliances. To do so, they used data from the NBA in the United States of America.

3 THE NATIONAL BEACH VOLLEYBALL COMPETITION OF THE NETHERLANDS

3.1 THE SETTING OF THE NATIONAL BEACH VOLLEYBALL COMPETITION

The National Beach Volleyball Competition in the Netherlands is organized by NeVoBo[1]. This is the Dutch federation for both indoor and beach volleyball. Within the National Beach Volleyball Competition in the Netherlands there are multiple levels. The highest level is called the 'eredivisie', the second highest level is called the first division, followed by the second and third division. This paper will use data for all female and male players that participated in at least one tournament in the 'eredivisie' from 2007 to 2015.

In the summer (the period from half of May until the end of August), the NeVoBo organizes every weekend an 'eredivisie' tournament somewhere in the Netherlands, each time in a different location. For every tournament female and male players have the opportunity to form teams of the same sex and choose their partners freely, which makes them free agents in economic terms. So, for every single tournament two players choose to form a team. Apart from that, players are not forced to participate in every single tournament. (NeVoBo, 2013)

The system described here, with players being able to freely form different teams for every tournament and not being obliged to participate in all tournaments, is enabled by an individual ranking points system. Every beach volleyball player has the opportunity to collect points by participating in a tournament of the National Beach Volleyball Competition in the Netherlands or in an international CEV- or FIVB-tournament[2] (NeVoBo, 2013). The number of points that a player gains depends on the outcomes of a tournament and also on the level of the tournament (whether it is 'eredivisie', first division, etc.). The individual points collected in the last year by

[1] NeVoBo is the abbreviation for 'De Nederlandse Volleybal Bond' which is Dutch for the Dutch Volleyball Federation.

[2] The European Volleyball Confederation (CEV) is the institution that is responsible for governing all the National Federations throughout Europe and is recognized as such by the Fédération Internationale de Volleyball (FIVB). (CEV, 2011)

the two players that form a team are summed up when the team subscribes for a tournament. These team ranking points are used to rank teams that subscribed and classify them in the right tournament level.

By doing so, all teams that subscribe are classified into the right level. The NeVoBo starts allocating teams, starting at the highest level ('eredivisie') and then moving downwards to the lower levels. Allocating teams to the 'eredivisie' is being done in a particular order. Firstly, all teams that ended on the eighth place or higher in the 'eredivisie' will be allowed to participate in the 'eredivisie' tournament in the following week. Secondly, the two best teams from the first division, will promote to the 'eredivisie' the next week. However, these two rules only apply if teams stay in the same formation. If partners decide to switch, the rules do not apply anymore, which results in more free spots for remaining teams. Lastly, the rest of the 'eredivisie' tournament will be completed by the teams that have the highest team ranking points until all sixteen male and sixteen female teams have been assigned. (NeVoBo, 2015)

The setting of the National Beach Volleyball competitions of the Netherlands as described above is precisely documented on their website[3]. The website contains data of all 'eredivisie' tournaments from 2006 till 2015. Using a data scraping technique, this data can be extracted from the website. This way three different datasets are formed. The first dataset contains information about all the 'eredivisie' tournaments in the given period defining, where and when the tournament took place, the teams that participated each tournament, the two players that formed a team, the ranking of the teams before a tournament started and the outcome of the teams in the tournaments. In total, this dataset has 2464 observations. A second dataset consists of player characteristics for every individual player that is included in the first dataset. These characteristics are gender, age, hometown, and (beach) volleyball club. This is data on player level with 419 observations. The website of the National Beach Volleyball Competition does not contain information about the length of the players, although this is expected to be very useful for this study. For this reason, the lengths of all 419 players are hand collected, by searching on different

[3] The website of the National Beach volleyball Competitions of the Netherlands: http://www.beachcompetitie.nl/.

websites[4]. The third dataset contains all Total Rank Points for every year of each beach volleyball player, with a total of 4,181 observations. This dataset is used to create a variable that represents the players' Total Rank Points of the season prior to a tournament.

3.2 THE MATCHING GAME WITHIN BEACH VOLLEYBALL

Since beach volleyball players are able to freely form a team with any other player from the same gender, the beach volleyball competition can be seen as a market where individuals form one-to-one two-sided matches. Therefore, a typical dataset for matching markets described by Fox (2006) can be used to find the characteristics maximizing the performance of matched beach volleyball players. In order to perform this empirical study, data from the 'eredivisie' of the National Beach Volleyball competition in the Netherlands from 2007 to 2015 has been studied.

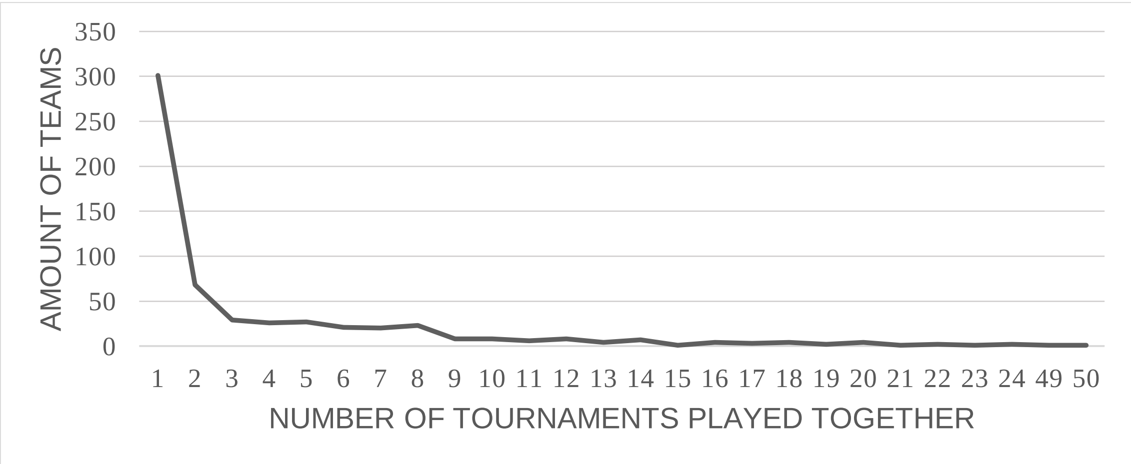
	mean	**2007**	**2008**	**2009**	**2010**	**2011**	**2012**	**2013**	**2014**	**2015**
86	9.6	9	12	12	13	8	9	7	8	8

Table 1: Number of tournaments per year

In the given period, 86 tournaments where played (table 1). The table shows that there are on average 10 beach volleyball tournaments a year. 390 beach volleyball players participated in at least one of these tournaments and formed 1,166 different teams in total over this period (table 2). The fact that the number of different teams is larger than the number of different players,

[4] The websites used to collect the length of all the beach volleyball players are the official website of the FIVB (http://www.fivb.org/EN/BeachVolleyball/PlayersDatabase.asp), the international website that contains Beach Volleyball Statistics (http://www.bvbinfo.com/player.asp), the Dutch website that contains a Volleyball Statistics (http://www.volleybalstatistieken.nl/), the websites of the players of the National Volleyball Team of the Netherlands (http://www.volleybal.nl/volleybal/oranje/oranje-heren and http://www.volleybal.nl/volleybal/oranje/oranje-dames), and some of the websites of Dutch volleyball clubs.

shows that beach volleyball players actually do switch beach partners during a year of tournaments. Graph 1 shows the number of tournaments that beach volleyball teams played together from 2007 to 2015. This graph clearly shows that beach volleyball players switch a lot between beach partners during the year. Almost 70% (Appendix 1) of the beach volleyball teams only played three times or less with the same beach volleyball partner and 90% (Appendix 1) of the teams nine times or less. Considering an average of 10 tournaments per year (table 1), only 10% (Appendix 1) of the teams actually played this amount of tournaments together. This could partly be explained by the number of tournaments beach volleyball players have played between 2007 and 2015. Almost 30% of the players only participated in 1 tournament and nearly 65% participated in 10 tournaments or less (Appendix 1). Still, graph 1 shows that beach volleyball players generally switch partners a lot. All these combinations of players in all these teams form an outstanding market to measure the relative importance of numerous player characteristics.



Graph 1: The number of tournaments that beach volleyball teams played together

The fact that approximately 30% of the players only played 1 tournament and roughly 65% played 10 tournaments or less (Appendix 1) shows that barely any player participated in all tournaments in the time period selected for this empirical study. This could be explained by four circumstances. First of all, the players could have been ill or injured and were therefore forced to skip a tournament. Secondly, the teams could have been promoted or relegated. On the one hand, there is the possibility of a team performing badly in a tournament, ending up at the bottom of the ranking. This causes the team to relegate to the first division. On the other hand, teams could perform so well that they get the chance to play internationally. The third circumstance is the fact that beach volleyball players are free to make the choice of participating in a tournament, which has made it a common practice for players to skip a few weeks, for instance due to holidays.

Another possibility is that beach volleyball players could participate in the highest level (‘eredivisie’) with one partner but play at a lower level (first division or lower) with another partner.

3.3 THE ELEMENTS THAT CAN INFLUENCE THE OUTCOME IN A TOURNAMENT

Variable	Type	Mean	Min	Max	Obs.	Missing (#Players)
Year	Continuous		2007	2015	4,464	0
PlayerID	Label		1	390	4,464	0
TeamID	Label		1	1,166	4,464	0
Outcome	Continuous	8	1	16.5	4,464	0
Ranking	Continuous	9	1	22	4,464	0
Female	Dummy	0.50	0	1	4,464	0
City	Categorical		1	136	4,460	4
Province	Categorical		1	11	4,441	22
Country	Categorical		1	5	4,459	5
Beach volleyball club	Categorical		1	33	2,836	200
Volleyball club	Categorical		1	94	2,538	146
Same city	Dummy	0.25	0	1	4,464	0
Same province	Dummy	0.52	0	1	4,464	0
Same country	Dummy	1.00	0	1	4,464	0
Same beach club	Dummy	0.28	0	1	4,464	0
Same indoor club	Dummy	0.12	0	1	4,464	0
Home city (P1)	Categorical	0.03	0	1	4,464	0
Home province (P1)	Categorical	0.20	0	1	4,464	0

Table 2: Descriptive statistics of the variables

It is possible to measure the success of a team by the outcome of a tournament. The outcome ranges between the first place and the seventeenth place. (Table 2) The outcome is sometimes rounded since games between low ended teams are skipped in a tournament. the Total Rank Points of the players, the number of times that the beach volleyball players played a tournament together, and some player and tournament characteristics are expected to mainly influence the outcome.

Variable	Female				Male			
	Mean	Min	Max	Obs.	Mean	Min	Max	Obs.
Total Rank Points	1,052	0	5,943	2,232	1,049	0	7,043	2,232
Ability	2,104	0	11,886	2,232	2,099	0	14,085	2,232
Experience	3	0	23	2,232	4	0	46	2,232
Age	24	11	43	2,232	26	15	50	2,232
Length	1.79	1.68	1.94	1,461	1.94	1.80	2.12	1,641
Diff. ability	0	-4323	4,323	2,232	0	-5,555	5,555	2,232
Diff. age	0	-27	27	2,232	0	-31	31	2,232
Diff. length	0	-0.26	0.26	1,172	0	-0.32	0.32	1,412

Table 3: Descriptive statistics of variables for females and males (all continuous variables).

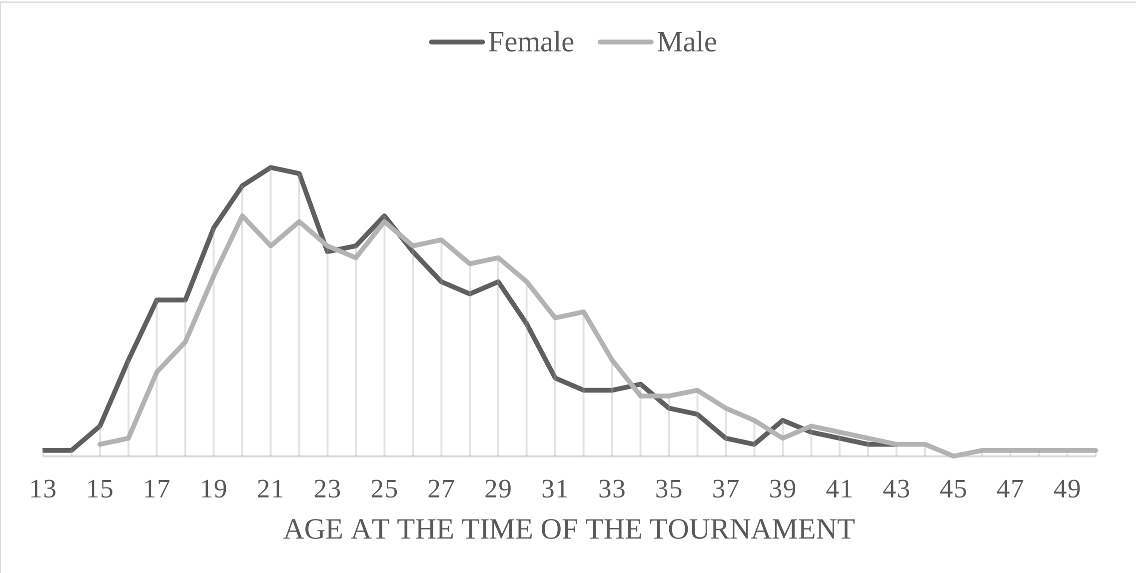
The Total Rank Points are expected to influence the outcome, since they are a measurement for earlier performance of the players. In this study, the Total Rank Points will be used in two ways. The Total Rank Points helps to determine the players' ability. The variable Ability presents the performance of a player in the previous *season* by summing up the Total Rank Points gained by the player in that last season.

To compute the variable team ability, the Total Rank Points of the two players in a team are summed up. The ability of the teams varies from 0 to 12,000 for females and from 0 to 14,000 for males. The average team ability is equal to 2,000 for both males and females (Table 3). This relatively low mean is expected, as 30% of the players only participated in one tournament and 60% joined no more than 10 tournaments. The relatively high maximum of 14,000 belongs to the beach volleyball players that also participate in international CEV- or FIVB-tournaments.

Secondly, the Total Rank Points help to rank the teams before the tournament starts in the current season. The performance of the players in the current season is expected to influence the outcome of the next tournament. This could be measured by the ranking of the teams before the tournament starts. The ranking is an ordinal scale based on the Total Rank Points that the two players of the teams gained in the last 365 days. So, while the variable Ability is computed by the Total Rank Points of the two players in a team earned in the previous *season*, the variable Ranking is computed by the Total Rank Points of the two players in a team in the previous 365 days. Since it is an ordinal scale, ranking also gives insights into the previous performance of the team compared to the opponent teams. This varies between the first place and the twenty-second place (Table 2). Both Total Rank Points in the previous year and ranking prior to the tournament are expected to have a positive effect on the outcome.

Moreover, the number of tournaments that the two players of a team participated in together, is a measure for the experience of the team. Team experience can hugely improve performance and is therefore expected to have a positive effect on the outcome. In this study the number of tournaments played together can be counted from 2007 onward. As a result, team experience before 2007 is not taken into account. The distribution of team experience is shown in graph 1. Table 3 contains descriptive statistics of team experience for female and male teams. Female teams have played together in the given period for a maximum of 23 times and male teams for a maximum of 46 times. The average number of tournaments played together is very low, respectively 2.5 and 4 for female and male teams. These averages support the statement made earlierthat beach volleyball players often switch teams. In 51% of the cases, players did not play any tournament together in the past (Appendix 1).

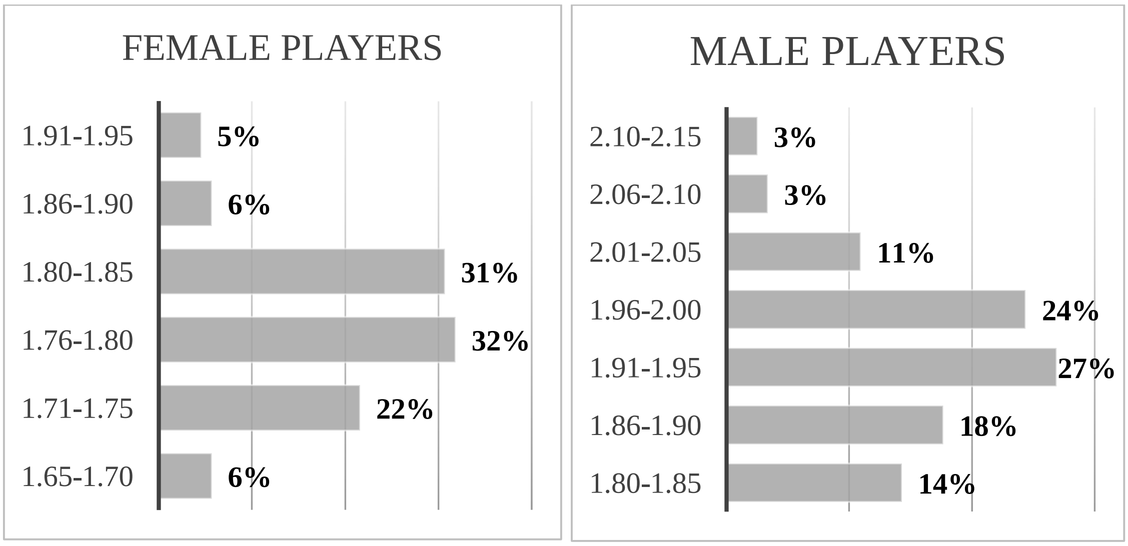
Further, there are some player characteristics that are expected to influence the outcome and should therefore be controlled for. These are a player's age, length, beach volleyball club, indoor volleyball club and the country, region and city the player comes from. All these characteristics are on an individual player-level. This study focuses on the characteristic gender, since differences are expected between female and male beach volleyball players. 191 males and 199 females formed teams in this period and exactly 50% of the outcomes are for female teams (2,232 observations) (table 2).



Graph 2: Number of beach volleyball players with a certain age at the time of the tournament.

Furthermore, age of the players is expected to influence the outcome of a tournament. The higher the age of the beach volleyball player, the more experience one is assumed to have. Therefore, a better outcome in a tournament is expected. Nevertheless, an older beach volleyball player may become less fit which could lead to a worse outcome. Graph 2 shows the age of the beach volleyball player at the time of the tournament. Both for female and male players, the distribution of the age is a little skewed to the right. This means that most players are somewhat younger and there is a longer tail representing older players. The second fact visible in this graph, shows that female players are on average younger than male players.

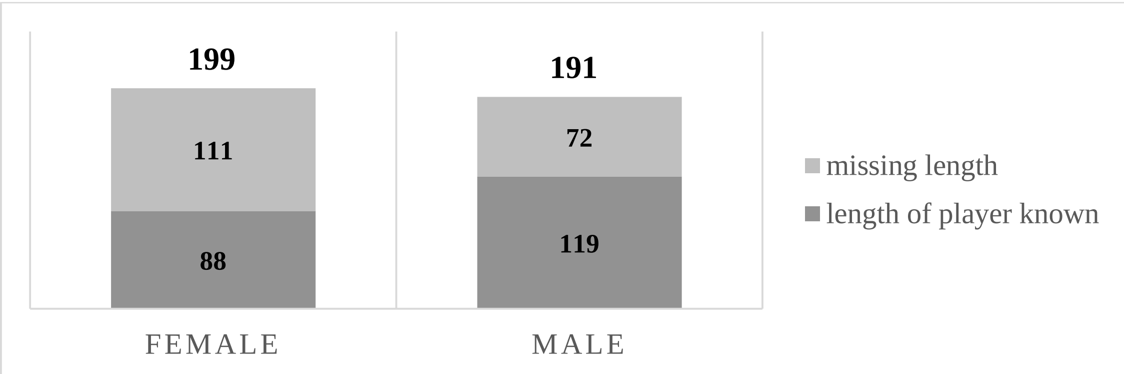
Moreover, length is expected to be very important in beach volleyball. This is reflected by the average length of beach volleyball players compared to the average length of the rest of the Dutch population. The average length of the female beach volleyball players is 1.79 metres and 1.94 metres for male beach volleyball players (table 3). This is more than ten centimetres longer than the average length of the Dutch population, which is 1.68 and 1.81 respectively for females and males in 2011 (Centraal Bureau voor de Statistiek, 2015). These lengths are almost exactly the same as the smallest lengths in the National Beach Volleyball Competition, namely 1.68 metres for female players and 1.80 for male players (table 3). The longest length in the National Beach Volleyball Competition for female players is 1.94 metres and 2.12 metres for male players (table 3).



Graph 3: Proportion of female and male players in the different length categories.

The distribution of the length of the female and male beach volleyball players is shown in graph 3. This graph displays for both male and female players an approximately normal distribution.

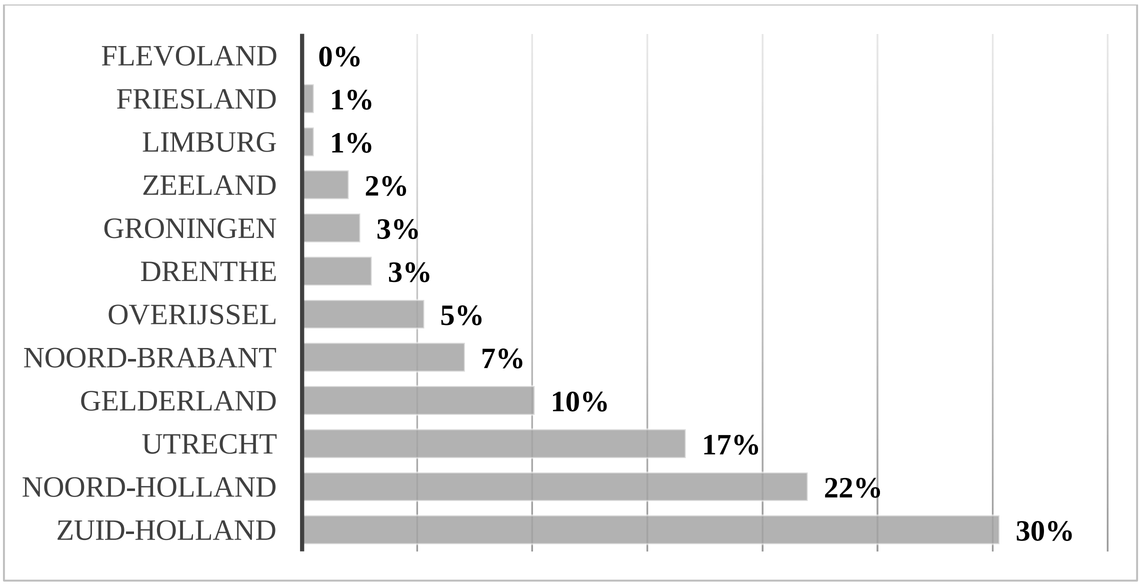
Since length seems to be so important, the expectation is that length will have a positive effect on the outcome in a tournament. The length is missing for many players, namely for 111 of the 199 female and 72 of the 191 male beach volleyball players (graph 4). Randomly missing lengths do not cause any problems in general. However, there is a chance that only the length of players with higher points is known. Appendix 2 shows that the average outcome in a tournament is better when the length is known and that the average Total Rank Points of the player are higher. Therefore, it is expected that the missing values of length are not random. The chance that a player's length is available online is assumed to be higher when the player performs better in tournaments.



Graph 4: the number of players which length is missing and which length is known.

Further, the place where a player lives, the beach and the indoor volleyball club can give information about how the beach partners know each other and can therefore partly explain the beach partner choice.

96% of the 390 beach volleyball players come from the Netherlands, only 17 players come from other countries (Belgium, Australia, Germany, and Spain) and 5 players' countries are unknown (Appendix 3).



Graph 5: The spread of the beach volleyball players over the provinces.

The beach volleyball players live in 136 different cities, only 4 players have an unknown city (table 2). From all players who come from the Netherlands and their city is not missing, the Dutch province they live in has been visualised in graph 5. The graph shows that the distribution of the players within the Netherlands is very unequal. Most players come from Zuid-Holland and Noord-Holland, which are the two provinces with the highest amount of inhabitants. Besides, Utrecht has a really high percentage of beach volleyball players. Especially when taking into account that this province comes on the fifth place of provinces with the highest amount of inhabitants. (Centraal Bureau voor de Statistiek, 2011) Only few beach volleyball players come from Drenthe, Groningen, Zeeland, Limburg and Friesland and nobody comes from Flevoland (graph 5).

Looking at the volleyball clubs, there are 94 different indoor volleyball clubs and 33 different beach volleyball clubs (table 2). For 146 beach volleyball players the indoor volleyball club is unknown and for 200 beach volleyball players the beach volleyball club is unknown (table 2). Remarkable is that there is far more missing data for beach volleyball clubs than for indoor volleyball clubs. There are three possible explanations for this.

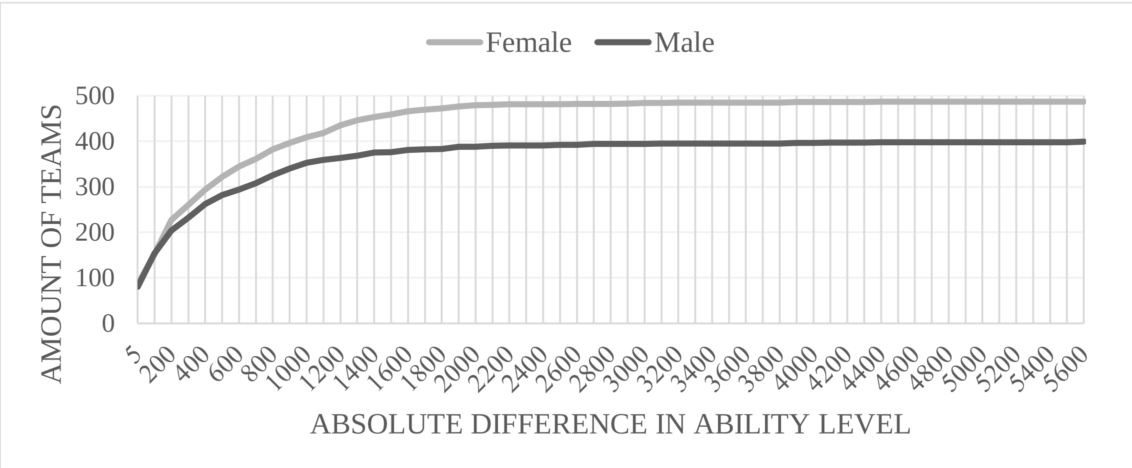
1. Many beach volleyball players practice indoor volleyball when it’s too cold for beach volleyball.
2. Most beach volleyball players started indoor volleyball and converted to beach volleyball. This explains why so many players mentioned playing in an indoor club rather than a beach club.

3. Many Beach Volleyball teams provide their own Beach Volleyball field and trainer. These teams do not join a Beach Volleyball Club.

Moreover, since beach volleyball is a very tactically game that can be influenced by many external factors, tournament characteristics captured by location are expected to influence the outcome. These tournament-level characteristics could be linked to the player-level characteristics. With that, it is possible to find out whether players participate in a tournament in their hometown. This is expected to have a positive effect on the outcome, since the players are well-known with all conditions of that location. For example, the sand can be very different on different locations and more importantly, the wind conditions can differ enormously . Wind is very influential in beach volleyball. Experience with the specific kind of wind on a location can have a huge impact on the performance in a tournament. Participating in a tournament in the player's hometown only occurs in 3% of the cases. Therefore, the paper will investigate the effect of home advantage. The expectation is that participating a tournament in the province where the player lives, has a positive effect on the outcome. This occurs in 20% of the cases (Table 2). Many players train in surrounding cities and not the exact city where they live in. Therefore, comparing provinces seems a much better measure than comparing cities.

3.4 Dispersion of the ability and characteristics of the team members

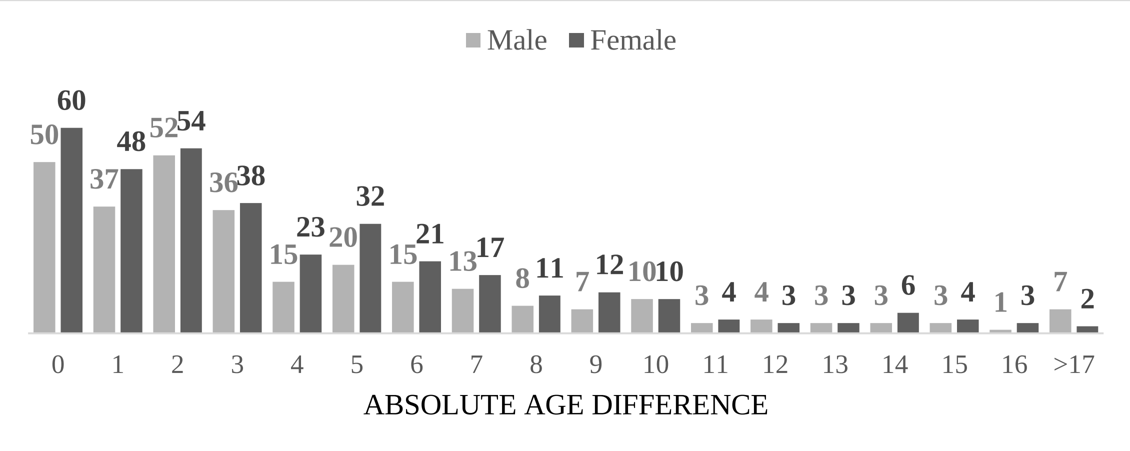
Beach volleyball players are more dispersed if their ability, age or length differ more from each other. This paper will examine the effect of the dispersion of a player's ability, length and age.



Graph 6: Cumulated number of teams within a certain ability difference.

First of all, the distribution of the differences in ability of all the teams is shown in graph 6. The higher a player's ability is, the more experienced that player is. The question is, whether the outcome will be higher if two high-ability players play together or if a high-ability player and a low-ability player forms a team together. To test this, the following hypothesis has been created:

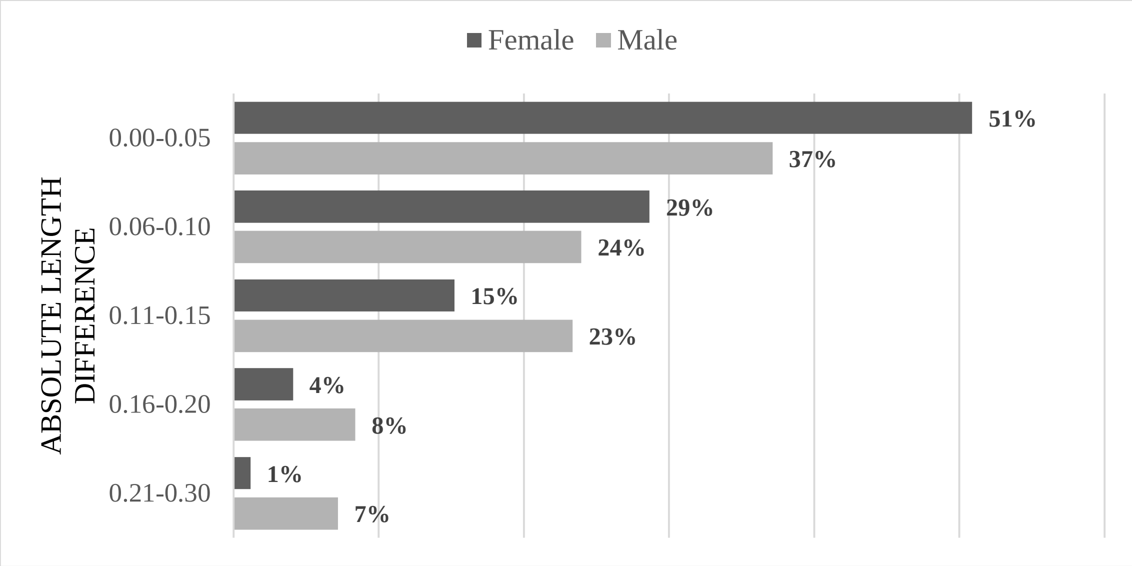
> *H1: dispersion of a player's ability will have a positive effect on the outcome in a tournament.*



Graph 8: Number of female and male teams within a certain age difference.

Secondly, because of the two contrasting expectations on age (on the one hand, an older player will have more experience, but, on the other hand, he might be less fit), expected is that the combination of an older and more experienced player with a younger and more fit player will be optimal. Thus, age dispersion is expected to have a positive effect on the outcome in a tournament, examining the second hypothesis:

> *H2: dispersion of the age of the players will have a positive effect on the outcome in a tournament.*



Graph 7: Proportion of female and male teams within a range of length difference.

Thirdly, the length of the two players that form a team are expected to deviate from each other. The length difference varies from -0.26 to 0.26 for females and -0.32 to 0.32 for males. In graph 7 this length difference will be divided in five groups with a range of five centimetres. It includes the percentages of female and male teams that are in a certain group. Most of the time one of the beach volleyball players is a little taller and the other one is somewhat shorter. The taller player is the player that goes to the volleyball net to set the block and the shorter player is a lot faster in the backcourt and is therefore able to save the ball from the sand. Length dispersion is therefore expected to have a positive effect on the outcome in a tournament. This will be tested with the third hypothesis, which is:

> *H3: dispersion of the length of the players will have a positive effect on the outcome in a tournament.*

However, the actual length of the players is expected to have a bigger effect on the performance, since two very small beach partners with a big length difference are expected to lose from two really big beach partners that do not differ much in length.

4 METHODOLOGY

4.1 THE DATASET

A single observation in the dataset represents the outcome in a tournament of two beach volleyball players that form a team in that particular tournament. Since the two players in a team are ordered alphabetically on the website, some players will always be player 1 and some players will always be player 2. Subtracting player characteristics of player 1 from player 2 will give positive or negative results depending on which player comes first. In order to prevent biased distributions, the dataset is doubled. Because of this, every observation will occur twice in the dataset, where player 1 and player 2 are reversed the second entry.

Lastly, all observations with missing values for outcomes will be removed. Since these observations are no-shows (mostly because of illness or injuries) of teams that did subscribe for the tournament. Moreover, all observations from the season of 2006 will be removed, since there is no information about the Total Rank Points of the players in 2005. Eventually this empirical study uses a dataset with data for ten beach volleyball seasons, from 2006 to 2015. This final dataset contains 4,464 observations, of which 2,232 are unique (table 2).

4.2 METHODS USED

To estimate the effect of dispersion of the characteristics of a beach volleyball team on the outcome in a tournament, several models are estimated. A primary model has been made in order to find all factors that influence the team's outcome in a tournament. When the primary model is set, the effect of dispersion of the characteristics can be measured. Which will be done by means of some models that compare the differences between players that form a team.

Since players switch partners so often, the data can be used to compare the differences and effect of dispersion by comparting one player to the different partners. To do so fixed effects are added to all models Player. So, all models only include the within variation and the variation between teams with completely different players will be ignored.

To establish these models, several variables will be transformed. For the original outcome and ranking variables '1' is the best score, since the winner gets the first place. This complicates the interpretation of the variable (e.g. the higher the outcome variables, the worse the beach volleyball players performed in a tournament). To simplify the interpretation of the coefficients, the highest outcome or ranking should be the best. Therefore, the variables will be transformed as in equation 1 (where y represents the outcome and ranking variables).

$$lny = ln\left(\frac{y_{max}-y+1}{y}\right) \quad (1)$$

Here, y_{max} is the maximum outcome or ranking in the tournament. So, the transformed variable expresses the outcome or ranking compared to the maximum outcome or ranking in the tournament that the team participates in.[5]

A similar transformation will be accomplished for the team ability variable, the team's experience together, team age and team length.[6] Likewise, these variables will be estimated relatively to the maximum of the teams in the tournament that they participate in (see equation 2, where x represents the variables).

$$lnx = \ln\left(\frac{x}{x_{max}-x+1}\right) \quad (2)$$

Because of the transformation in equation 2, the models will examine, for example, the effect of an increase in the team's length compared to the tallest team of the tournament. Contrary to just testing the effect of a centimetre difference in length of a team, which is not expected to have a major impact on the outcome as it is a minimal difference.

To derive the differences in ability, age and length of the two players that form a team, the ability, age or length of player 1 will be subtracted from the ability, age or length of player 2. Likewise, these variables will be transformed in line with equation 2. In addition, the variable age is originally constant over time. By subtracting the number of years since the tournament was played from the age of the player, the variable will fluctuate over the years.

[5] This is comparable to the variables that Szymanski (2000) created. This article compared the variables to the average of that variable.

[6] The summed age and summed length of a team is derived by adding op the age or length of the two players that forming a team.

A few models will be made that contain the factors that influence the outcome in a tournament. This will be attained by an Ordinary Least Squares (OLS) method with robust standard errors. These models are based on the following equation:

$$lnoutcome = \beta_0 + \beta_1 * \ln ranking + \beta_2 * \ln team\ ability + \beta_3 * \ln \text{team experience}^2 + \beta_4 * home\ advantage + \beta_5 * \ln team\ age + \beta_6 * female * \ln team\ age + \beta_7 * \ln team\ length + \beta_8 * female * \ln team\ length + \beta_9 * \ln ability\ difference + \beta_{10} * \ln age\ difference + \beta_{11} * \ln length\ difference + \varepsilon$$

The natural logarithm of the outcome relative to the maximum outcome in that particular tournament is the dependent variable which is called *ln outcome*. The explanatory variables are the following:

- the natural logarithm of ranking before a tournament, relative to the maximum ranking in that particular tournament (*ln ranking*),
- the natural logarithm of the summed Total Rank Points, relative to the maximum Total Rank Points in the tournament (*ln team ability*),
- the natural logarithm of the square of the number of tournaments that the two players in a team participated in together, relative to the maximum team experience in the tournament (*ln team experience*^{2}),
- a dummy which is equal to 1 if the player lives in the same province as the location of the tournament and 0 if otherwise (*home advantage*),
- a dummy with the value 1 if the beach volleyball players are females and 0 if otherwise (*female*),
- the natural logarithm of the summed age of the players, relative to the maximum summed age in the tournament (*ln team age*),
- an interaction variable between gender and the summed age variable (*female * ln team age*),
- the natural logarithm of the summed length of the players relative to the maximum summed length in the tournament (*ln team length*),
- and an interaction variable between gender and the summed length variable (*female * ln team length*).

Furthermore, there are three explanatory variables that represent dispersion of ability, age and length, namely:

- the natural logarithm of the players' Total Rank Point difference relative to the maximum Total Rank Point difference in the tournament (*ln ability difference*),
- the natural logarithm of the players' age difference relative to the maximum age difference in the tournament (*ln age difference*),
- and the natural logarithm of the players' length difference relative to the maximum length difference in the tournament (*ln length difference*).

The correlation matrix (Appendix 4) does not show any unexpected high correlations. The dependent variable ln outcome particularly has a high correlation with the variables ln ranking and ln team ability.

5 RESULTS

The primary models are discussed first. These OLS models test the effects of different variables on the outcome in a tournament. Hereafter, the hypotheses will be tested by including the dispersion variables (difference in ability, age and length) to the variables from the primary models that have a significant effect on the outcome in a tournament.

5.1 THE PRIMARY MODELS

	1	2	3	4	5	6
ln ranking	0.58***	0.58***	0.58***	0.58***	0.58***	0.53***
	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)
ln team ability	0.18***		0.18***	0.19***	0.18***	0.17***
	(0.02)		(0.02)	(0.02)	(0.02)	(0.02)
ln team ability^2		0.09***				
		(0.01)				
female * ln team ability			0.01			
			(0.03)			
ln team experience^2				-0.01		
				(0.01)		
home advantage					0.14**	0.13**
					(0.04)	(0.04)
length known						0.48***
						(0.08)
Constant	-0.17***	-0.17***	-0.17***	-0.19***	-0.19***	-0.47***
	(0.03)	(0.03)	(0.03)	(0.06)	(0.03)	(0.05)
Observations	4,464	4,464	4,464	4,464	4,464	4,464
R^2	0.4591	0.4591	0.4592	0.4593	0.4604	0.4803
Adjusted R^2	0.4589	0.4589	0.4588	0.4590	0.4600	0.4799

Table 4: The effect on the variable ln outcome (: significant at 5%, **: significant at 1%, ***: significant at 0.1%).*

The first model (table 4) contains the logarithm of the variables ranking and team ability. Both variables have a positive and significant effect on the logarithm of the outcome in a tournament at a significance level of 1%. If the ranking will be 1% higher, the outcome in a tournament will be 0.6% higher and if the ability of the team gets 1% higher, the outcome in a tournament will be

0.2% higher, ceteris paribus. The results of model 1 confirms the expectations. The adjusted R-squared of this model is 0.46, which means that the outcome in a tournament could be explained for 46%, by:

- the ranking at the start of the tournament (relative to the maximum ranking)
- the ability of the team (relative to the team with the highest ability in the tournament).

The better the team performed in the past, the larger the chance that they will end high in the tournament.

The logarithm of the squared team ability of the team has replaced the team ability variable in model 2 (table 4). The variable has a significant coefficient at a significance level of 1%, but the adjusted R-squared has not increased.

Model 3 (table 4) shows that the ability level has no different effect for female and male players. Since the interaction variable between gender and team ability does not have a significant effect.

In conclusion, comparing these three models, model 1 will be preferred over model 2 and 3.

The effect of the team's experience has been tested in model 4 (table 4). No significant relation between the number of times that the players who forms a team have played together in the past and the outcome in a tournament are found. Therefore, the experience variable will not be added in the models from now onward.

In model 5 (table 4) a dummy is added that denotes home advantage. The dummy is equal to one if a player lives in the same province as the location of the tournament and zero otherwise. This gives a significant outcome at a significance level of 1%. Playing a tournament in the same province as the province where a player lives, compared to playing in another province, increases the outcome in the tournament by 0.1%, ceteris paribus. This confirms the expectation that players who are familiar with all conditions of that location will have a higher outcome in a tournament (i.e. home advantage does exist in beach volleyball). The other variables remain the same as in model 1. As well as the adjusted R-squared, which is still equal to 0.46.

The length is only known for part of the players (44% of the female players and 62% of the male players). In the ideal situation, the group with players whose length is known, is a random selection. Since this would make sure that there is no difference in outcome between the group of players whose length is known and the group of players whose length is unknown. To test whether the two groups are randomly selected, a dummy is added in model 6 (table 4). The dummy is equal to one if a player's length is known and zero otherwise. The significantly positive coefficient shows that a known length of a beach volleyball player, compared to an unknown length, increases the outcome in a tournament, ceteris paribus. This confirms the expectation that lengths are more likely to be known for better Beach Volleyball players and unknown for less performing players. The remaining part of the model remains the same as model 5. The adjusted R-squared rises to 0.48.

In conclusion, when comparing models 5 and 6 to the previous models, they are preferred. Model 6 has the highest adjusted R-squared, but the result could be biased because of the dummy variable that is equal to 1 if the length is known. Since a lot of the lengths of players are missing, this could influence the results.

	7	**8**	**9**	**10**	**11**
ln ranking	0.58^{***}	0.53^{***}	0.53^{***}	0.52^{***}	0.52^{***}
	(0.03)	(0.03)	(0.04)	(0.03)	(0.03)
ln team ability	0.18^{***}	0.17^{***}	0.17^{***}	0.17^{***}	0.17^{***}
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
home advantage		0.13^{**}	0.13^{**}	0.13^{**}	0.12^{**}
		(0.04)	(0.04)	(0.04)	(0.04)
length known		0.49^{***}	0.50^{***}	0.51^{***}	0.50^{***}
		(0.08)	(0.08)	(0.08)	(0.08)
ln team age	-0.03	0.03	0.02		-0.08^{*}
	(0.03)	(0.03)	(0.04)		(0.04)
female * ln team age			0.03		
			(0.05)		
ln team age^2				0.16^{*}	0.33^{**}
				(0.08)	(0.10)
Constant	-0.14^{**}	-0.51^{***}	-0.51^{***}	-1.76^{**}	-2.96^{***}
	(0.04)	(0.05)	(0.05)	(0.59)	(0.79)
Observations	4,464	4,464	4,464	4,464	4,464
R^2	0.4595	0.4806	0.4808	0.4819	0.4827
Adjusted R^2	0.4592	0.4800	0.4801	0.4813	0.4820

Table 5: The effect on the variable ln outcome (: 5%, **: 1%, ***: 0.1%).*

Adding the logarithm of the team age of the team to model 1 and 6 does not give any significant results for this variable (see respectively model 7 and 8 in table 5). Moreover, including an interaction variable that multiplies the gender dummy with the team age of the team does not give any significant results for the team age either (model 9 in table 5). However, the logarithm of the squared team age of the team in model 10 (table 5) does give a significant coefficient at the 5% significance level. When the squared sum of the two players' age increases with 1%, the outcome in a tournament increases with 0.2%, ceteris paribus. So, there seems to be a positive non-linear relation between the outcome in a tournament and experience of the players because of their team age. The other variables remain significant with the same signs occurring.

Model 11 (table 5) includes both the logarithm of the team's summed age and the logarithm of the squared team's team age. When the sum of the two players' age increases with 1%, the outcome in a tournament decreases with 0.1% and when the squared sum of the two players' age increases with 1%, the outcome in a tournament increases with 0.3%, ceteris paribus. For both model 10 and 11, the adjusted R-squared increases to 0.48.

In conclusion, there seems to be a non-linear relation between the team's summed age and the outcome in a tournament. Since adding the square of the team's age gives significant effects on the outcome in a tournament. Therefore, model 11 is still preferred, because it has the highest adjusted R-squared so far.

	12	13	14	15	16	17
ln ranking	0.46^{***}	0.46^{***}	0.46^{***}	0.46^{***}	0.46^{***}	0.46^{***}
	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)
ln team ability	0.20^{***}	0.20^{***}	0.20^{***}	0.20^{***}	0.20^{***}	0.20^{***}
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
home advantage	0.14^{**}	0.15^{**}	0.15^{**}	0.14^{*}	0.14^{*}	0.14^{*}
	(0.06)	(0.06)	(0.06)	(0.05)	(0.05)	(0.06)
ln team length	-0.03	-0.02			0.01	
	(0.03)	(0.03)			(0.05)	
female * ln team length		0.04			-0.08	
		(0.03)			(0.06)	
ln team length^2			-0.47	0.59	0.65	
			(0.47)	(0.82)	(1.07)	
female * ln team length^2				0.09	0.17	0.05
				(0.07)	(0.09)	(0.04)
Constant	0.10	0.03	1.26	-1.63	-1.83	-0.04
	(0.11)	(0.11)	(1.26)	(2.25)	(2.81)	(0.07)
Observations	2584	2584	2584	2584	2584	2584
R^2	0.4089	0.4098	0.4094	0.4110	0.4116	0.4106
Adjusted R^2	0.4079	0.4086	0.4085	0.4099	0.4100	0.4097

Table 6: The effect on the variable ln outcome (: 5%, **: 1%, ***: 0.1%).*

In models 12 to 17 (table 6) the logarithm of the teams length, the logarithm of the squared team length and the interactions of these variables with the gender dummy are added to model 4. None of these length variables show a significant effect. This implies that the length of a beach volleyball player does not have any significant effect on the outcome in a tournament. A reason for this could be the fact that all individuals (from whom the length is known) are taller than the average citizen of the Netherlands. In order to play in the highest beach volleyball tournaments of the Netherlands it is a requirement to be taller than average, but the variation between all these tall players does not show any effects on the outcome in a tournament.

To conclude: model 11 is the best model. The model has the highest adjusted R-squared of 0.48. Notably, initial model 1 already has an adjusted R-squared of 0.46, indicating that most of model 11 is explained by the two variables used in model 1: Ranking and team ability of the team. The adjusted R-squared only rises to 0.48 in model 11 compared to 0.46 in model 1 by adding home advantage, known length, the natural logarithm of the teams summed age and the natural logarithm of the squared teams summed age.

5.2 DISPERSION OF THE CHARACTERISTICS OF THE BEACH VOLLEYBALL PLAYERS

	18	19	20
ln ranking	0.56***	0.57***	0.52***
	(0.04)	(0.03)	(0.03)
ln team ability	0.19***	0.19***	0.17***
	(0.02)	(0.02)	(0.02)
ln ability diff.	0.03**		
	(0.01)		
ln ability diff. 2		0.02***	0.01***
		(0.00)	(0.00)
home advantage			0.12**
			(0.04)
length known			0.48***
			(0.08)
ln team age			-0.08*
			(0.04)
ln team age2			0.33**
			(0.10)
Constant	-0.08	-0.07	-2.86***
	(0.05)	(0.04)	(0.79)
Observations	2580	4464	4464
R^2	0.4614	0.4643	0.4853
Adjusted R^2	0.4607	0.4639	0.4845

Table 7: The effect on the variable ln outcome (: significant at 5%, **: significant at 1%, ***: significant at 0.1%).*

The dispersion of the ability can be measured by adding the logarithm of the ability difference of the two players to model 1. Model 18 (table 7) shows that this variable has a significant coefficient at a 1% significance level. When the ability difference is 1% higher, the outcome in a tournament will be 0.03% higher, ceteris paribus. In model 19 (table 7) the logarithm of the squared difference of the ability has approximately the same coefficient with the same significance. These two models are in line with the expectations and hypothesis 1. Dispersion of the ability of the beach volleyball players within a team has a significantly positive effect on the outcome.

In model 20 (table 7) the logarithm of the squared difference of the ability has been added to model 11, since this was the best model so far. Now, the adjusted R-squared slightly rises (0.48) and model 20 is preferred over model 11.

The positive effect of dispersion of players' abilities could be explained by the fact that the low-ability player can gain and learn from the ability of the high-ability player. The result indicates that the low-ability will perform better when coupled to a high-ability player, leading to better outcomes in the tournament.

The dispersion of the age and the length of the beach volleyball players does not give any significant results.

	21	**22**	**23**	**24**
ln ranking	0.59***	0.58***	0.58***	0.52***
	(0.04)	(0.03)	(0.03)	(0.03)
ln team ability	0.18***	0.19***	0.19***	0.17***
	(0.02)	(0.02)	(0.02)	(0.02)
ln team age	-0.11*	-0.12***	-0.12***	-0.08*
	(0.05)	(0.04)	(0.04)	(0.04)
ln team age^2	0.17	0.26**	0.26**	0.31**
	(0.13)	(0.10)	(0.10)	(0.11)
ln age diff.	0.03			
	(0.03)			
ln age diff.^{2}		0.00	0.00	0.01
		(0.01)	(0.01)	(0.01)
female * ln age diff.^{2}			0.01	
			(0.02)	
home advantage				0.12**
				(0.04)
length known				0.51***
				(0.08)
Constant	-1.36	-2.10**	-2.07**	-2.80***
	(0.99)	(0.76)	(0.73)	(0.82)
Observations	2499	4464	4464	4464
R^2	0.4728	0.4610	0.4610	0.4829
Adjusted R^2	0.4717	0.4604	0.4603	0.4821

Table 8: The effect on the variable ln outcome (: significant at 5%, **: significant at 1%, ***: significant at 0.1%).*

In models 21 to 24 (table 8), the logarithm of the players' difference in age, the squared difference of the age and the interaction of the variables with the gender dummy are included. Therefore, hypothesis 2 is rejected. Dispersion of the age of the two beach volleyball players does not have any effect on the outcome.

Model 11 showed that the age of the two players in a team does have a significant positive effect on the outcomes in a tournament. This indicates that the older the two players within a team, the higher their outcomes in a tournament, keeping all other factors constant. Which means that experience because of age is more needed to perform better than being young and fit.

	25	**26**	**27**	**28**	**29**	**30**
ln ranking	0.46***	0.46***	0.46***	0.46***	0.46***	0.46***
	(0.06)	(0.04)	(0.06)	(0.04)	(0.06)	(0.06)
ln team ability	0.20***	0.20***	0.20***	0.20***	0.20***	0.20***
	(0.03)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
home advantage	0.20**	0.15**	0.20**	0.13*	0.19**	0.19**
	(0.07)	(0.06)	(0.07)	(0.05)	(0.07)	(0.07)
ln team length	-0.02		-0.01		0.02	0.02
	(0.04)		(0.04)		(0.07)	(0.07)
ln team length^2		-0.45		0.69	0.36	0.36
		(0.49)		(0.82)	(1.53)	(1.53)
female * ln team length			0.03		-0.05	-0.05
			(0.05)		(0.09)	(0.09)
female * ln team length^2				0.05	0.05	0.05
				(0.07)	(0.13)	(0.13)
ln length diff.	0.01		0.03		0.06	0.06
	(0.03)		(0.04)		(0.12)	(0.12)
ln length diff.^{2}		-0.01		0.01	-0.07	-0.07
		(0.04)		(0.05)	(0.20)	(0.20)
female * ln length diff.			-0.03		0.13	0.13
			(0.06)		(0.18)	(0.18)
female * ln length diff.^{2}				-0.09	-0.28	-0.28
				(0.08)	(0.30)	(0.30)
Constant	0.07	1.20	-0.00	-1.90	-1.16	-1.16
	(0.14)	(1.32)	(0.15)	(2.24)	(3.98)	(3.98)
Observations	1339	2584	1339	2584	1339	1339
R^2	0.4080	0.4094	0.4092	0.4115	0.4127	0.4127
Adjusted R^2	0.4058	0.4083	0.4061	0.4099	0.4079	0.4079

Table 9: The effect on the variable ln outcome (: 5%, **: 1%, ***: 0.1%).*

In models 25 to 30 the logarithm of the players' difference in length, the squared difference of the players' length, the square of these variables and the interaction of these variables with the gender dummy do not give a significant outcome. The expectation that a longer player and a shorter player in one team will improve the outcome in a tournament (hypothesis 3) is rejected. Length difference of the two players has no significant effect. As mentioned before, in order to play in the highest beach volleyball tournaments of the Netherlands it is a requirement to be taller than average. Since all players are so tall and length differences are so minimal, the variation in length difference between players who form a team do not show any effects on the outcome in a tournament.

6 CONCLUSION

Team composition theory prescribes that dispersion, i.e. the heterogeneity of workers, within a team concerning workers' ability and skill levels enhances team performance. Elaborating on previous research, this study examines the effect of the dispersion of abilities, ages and lengths of pairs of volleyball players, forming a team together, on the performance of the team.

A unique dataset of the National Beach Volleyball Competition allows researching the hypotheses that dispersion of players' abilities, ages and lengths will all have positive effects on the outcomes, i.e. the position of the team's ranking, in a tournament. Using actual sports data providing a large sample size allows to test the hypotheses effectively.

The empirical results provide evidence for the first hypothesis: dispersion of players' abilities has a significant effect on the performance of the beach volleyball team, measured by the outcomes in tournaments. On the one hand, these findings are in line with previous research by Nickerson & Owan (2003) about heterogeneity of workers' abilities in the workplace having a positive effect on performance. On the other hand, these results deviate from prior research in which team incentives drive high-ability workers into forming teams with each other. This was found by Iranzo et al. (2008), who stated that it is optimal to match production and non-production workers with approximately the same skill level.

In practice, the findings suggest that a high-ability player will not maximize the teams' outcomes by teaming up with a player who also qualified as a high-ability player. The high-ability player is likely to have better teams' outcomes in the current season, if he chooses to team up with a player qualified as low-ability in the previous season. Adversely, a low-ability player in the previous season would maximize the teams' outcomes in the current season, if he teams up with a player qualified as high-ability is the previous season.

The positive effect of dispersion of players' abilities could be explained by the fact that the low-ability player can gain and learn from the ability of the high-ability player. The result indicates that the low-ability will perform better when coupled to a high-ability player, leading to better outcomes in the tournament.

There are no significant effects of the dispersion of the ages of players in a team on the outcomes of the tournaments. It was expected that in an optimal situation a combination of an older, experienced player and a younger, fit player would lead to higher outcomes in tournaments. The results do not support this hypothesis.

The total age of the two players in a team does have a significant effect on the outcomes in a tournament. This indicates that the older the two players within a team are, the higher their outcomes in a tournament, keeping all other factors constant. A possible explanation could be the fact that beach volleyball is a very tactical game. Therefore, experience is more needed to perform better than being young and fit. This is also confirmed by the positive effect when a beach volleyball team has home advantage. Home advantage ensures that the players are well-known with all conditions of that location (i.e. wind). In conclusion, experience in the sense of both age and home advantage lead to a better outcome in a tournament.

There is also no significant effect of length dispersion of players in a team on the outcomes of the tournament. It was expected that in an optimal situation the taller player is the player that goes to the volleyball net to set his block and the shorter player is a lot faster in the backcourt and is therefore able to save the ball from the sand. This is in fact usually the division of beach volleyball players in practice. The results do not support this hypothesis.

A possible explanation for not finding a significant effect of length dispersion on the teams' outcomes, might be the fact that all individuals (from whom the length is known) are taller than the average citizen of the Netherlands. Since the other length variables also show insignificant results, length does not improve the performance. This implies that a minimum length is required to play professional beach volleyball. The variation between all these tall players does not show any effect on the outcome in a tournament.

To conclude, the study indicates that teams of two beach volleyball players consisting out of a high-ability and a low-ability player will have higher outcomes in tournaments. Moreover, more experienced players, because of age and home advantage, will also perform better in tournaments. Therefore, dispersion of abilities in the beach volleyball team consisting of more experienced players could make an average team into a winning team.

LIMITATIONS

The first limitation of this paper is the fact that the length is only known from part of the beach volleyball players. The results showed that the group of players whose length is known is not a random sample of the dataset. This implies that there is a selection bias. The only way to improve this is to change the collection method of the length variable. The most valid method will be measuring all the beach volleyball players. This is very labour intensive. Therefore, another, slightly less valid, approach is making a survey and ask all players for their length (e.g. by e-mail).

Another limitation is the fact that variables like city, province, country, volleyball club and beach volleyball club are scraped from a website and therefore fixed over time. Of course, these player characteristics can change over time. Further research can improve these variables by finding out for the whole period of time the exact cities and clubs where the players lived and played volleyball. For the variable age, which was originally fixed over time, the problem was solved. By subtracting the number of years since the tournament was played from the age of the player, the variable fluctuates over the years.

7 References

Alchian, A. A., & Demsetz, H. (1972). Production, information costs, and economic organization. *The American economic review*, 777-795.

Becker, G. S. (1973). A theory of marriage: Part I. *The Journal of Political Economy*, 813-846.

Berri, D. J., & Jewell, R. T. (2004). Wage Inequality and Firm Performance: Professional Basketball's Natural Experiment. *Atlantic Economic Journal*, *32*, 130–139.

Bloom, M. (1999). The performance effects of pay dispersion on individuals and organizations. *Academy of Management Journal*, *42*(1), 25-40.

Borland, J., & Lye, J. (1996). Matching and mobility in the market for australian rules football coaches. Industrial & Labor Relations Review, 50(1), 143-158.

CEV. (2011). About the CEV. Retrieved from CEV: http://www.cev.lu/CEV-Area/cev.aspx

Centraal Bureau van de Statistiek. (2011, February 22). Bevolkingscijfers per provincie op 1 januari 2011. *STAATSCOURANT*, 3346, 1. Retrieved from Overheid.nl: https://zoek.officielebekendmakingen.nl/stcrt-2011-3346.pdf

Centraal Bureau voor de Statistiek. (2015, July 24). Lengte en gewicht van personen, ondergewicht en overgewicht; vanaf 1981. Retrieved from CBS: http://statline.cbs.nl/Statweb/publication/?VW=T&DM=SLNL&PA=81565NED&D1=2-6&D2=1-2&D3=5&D4=0&D5=0-30&HD=151118-1352&HDR=G1,T&STB=G2,G3,G4

Coates, D., Frick, B., & Jewell, T. (2014). Superstar Salaries and Soccer Success The Impact of Designated Players in Major League Soccer. Journal of Sports Economics, 1527002514547297.

Choo, E., & Siow, A. (2006). Who marries whom and why. Journal of political Economy, 114(1), 175-201.

d'Addona, S., & Kind, A. (2012). Forced Manager Turnovers in English Soccer Leagues: A Long-Term Perspective. Journal of Sports Economics, 1527002512447803.

Depken, C. A. (2000). Wage disparity and team productivity: evidence from major league baseball. Economics letters, 67(1), 87-92.

Dupuy, A., & Galichon, A. (2014). Personality traits and the marriage market. Journal of Political Economy, 122(6), 1271-1319.

Fox, J. T., & Santiago, D. (2007). A toolkit for matching maximum score estimation and point and set identified subsampling inference.

Fox, J. T. 2008. Estimating matching games with transfers. Working paper, University of Chicago, Chicago.

Haskel, J., Hawkes, D. D., & Pereira, S. C. (2005). Skills, human capital and the plant productivity gap: UK evidence from matched plant, worker and workforce data.

Hellerstein, J. K., & Neumark, D. (2007). Production function and wage equation estimation with heterogeneous labor: Evidence from a new matched employer-employee data set. In Hard-to-measure goods and services: Essays in honor of Zvi Griliches (pp. 31-71). University of Chicago Press.

Iranzo, S., Schivardi, F., & Tosetti, E. (2008). Skill Dispersion and Firm Productivity: An Analysis with Employer-Employee Matched Data. Journal of Labor Economics, 26(2), 247-285.

Katayama, H., & Nuch, H. (2011). A game-level analysis of salary dispersion and team performance in the national basketball association. *Applied Economics, 43*, 1193–1207.

Katz, M. L., & Shapiro, C. (1994). Systems competition and network effects. The Journal of Economic Perspectives, 8(2), 93-115.

Levine, D.I., 1991. Cohesiveness, productivity and wage dispersion. Journal of Economic Behavior and Organization 15, 237–255.

NeVoBo. (2013, December). Beachvolleybal Competitie. In Wedstrijdreglement: http://www.volleybal.nl/data/bestanden/competitie/naslag/reglementen-en-regelgeving/nationaal/reglementen/03-wr-hoofdstuk-6-beachvolleybal-competitie-december-2013.pdf

NeVoBo. (2015). Deelnemen aan de beachcompetitie. Retrieved from Beachcompetitie: http://www.beachcompetitie.nl/naslag/deelname

NeVoBo. (2015). Werkwijze indeling. Retrieved from Beachcompetitie: http://www.beachcompetitie.nl/naslag/indeling

Ramaswamy, R., Rowthorn, R.E., 1991. Efficiency wages and wage dispersion. Economica 58, 501–514.

Scully, G.W., 1989. The Business of Major League Baseball, The University of Chicago Press, Chicago, IL.

Szymanski, S. (2000). A market test for discrimination in the English professional soccer leagues. Journal of Political Economy, 108(3), 590-603.

Yang, Y., Shi, M., & Goldfarb, A. (2009). Estimating the value of brand alliances in professional team sports. Marketing Science, 28(6), 1095-1111.

8 APPENDIX

Appendix 1: Number of tournaments played by beach volleyball players and teams.

Tournaments played	Players			Teams		
	Amount	Percent	Cumulative	Amount	Percent	Cum.
1	107	27.4%	27.4%	301	51.7%	51.7%
2	22	5.6%	33.1%	68	11.7%	63.4%
3	19	4.9%	37.9%	29	5.0%	68.4%
4	17	4.4%	42.3%	26	4.5%	72.9%
5	21	5.4%	47.7%	27	4.6%	77.5%
6	13	3.3%	51.0%	21	3.6%	81.1%
7	13	3.3%	54.4%	20	3.4%	84.5%
8	14	3.6%	57.9%	23	4.0%	88.5%
9	10	2.6%	60.5%	8	1.4%	89.9%
10	14	3.6%	64.1%	8	1.4%	91.2%
11	7	1.8%	65.9%	6	1.0%	92.3%
12	6	1.5%	67.4%	8	1.4%	93.6%
13	8	2.1%	69.5%	4	0.7%	94.3%
14	3	0.8%	70.3%	7	1.2%	95.5%
15	6	1.5%	71.8%	1	0.2%	95.7%
16	5	1.3%	73.1%	4	0.7%	96.4%
17	9	2.3%	75.4%	3	0.5%	96.9%
18	8	2.1%	77.4%	4	0.7%	97.6%
19	3	0.8%	78.2%	2	0.3%	97.9%
20	5	1.3%	79.5%	4	0.7%	98.6%
21-30	42	10.8%	90.3%	6	1.0%	99.7%
31-60	38	9.7%	100.0%	2	0.3%	100.0%
Total	390	100%		582	100%	

Appendix 2: Differences in tournament outcome and Total Rank Points when length of the beach volleyball player is known or unknown.

Player 1	Length known				Length unknown			
	Mean	Min	Max	Cases	Mean	Min	Max	Cases
Outcome	7	1	17	3,102	10	1	17	1,362
Total Rank Points	1,160	0	7,043	3,102	801	0	1,812	1,362

Appendix 3: Countries that the beach volleyball players come from.

Country	Cases	Percent	Cum.
Netherlands	368	94.4%	94.4%
Belgium	7	1.8%	96.2%
Australia	5	1.3%	97.4%
Germany	3	0.8%	98.2%
Spain	2	0.5%	98.7%
Unknown	5	1.3%	100%
Total	390	100.0%	

Appendix 4: Correlation table.

		1	2	3	4	5	6	7
1	ln outcome	1.00						
2	ln ranking	0.56	1.00					
3	ln team ability	0.48	0.34	1.00				
4	ln team experience	0.15	0.17	0.21	1.00			
5	home advantage	0.08	0.07	0.03	-0.00	1.00		
6	ln team age	0.12	0.10	0.08	0.09	-0.07	1.00	
7	ln team length	0.00	0.00	0.07	0.09	-0.01	0.13	1.00