# Color Psychology in Football: The Effect of Shirt Color on a Team’s Performance in the Dutch Eredivisie

**Toby Tan 403228**

*Student Behavioral Economics, Erasmus University Rotterdam*

Supervisor: Yu Gao

## Abstract

Color psychology is the behavioral term for the phenomenon that colors influence people’s decisions and behavior. In the literature of sports, color psychology is a hot item because several authors have found opposite results in declaring the effects of shirt color. Attrill, Hill & Barton (2008) show a positive long-term effect of red colored shirts in the English Premier League, while Kocher & Sutter (2008) find no effects of shirt color in the German Bundesliga. This paper will contribute to the existing discussion and will provide an analysis on the shirt colors in the Dutch Eredivisie using several variants of the Propensity Score Matching method designed by Rosenbaum and Rubin (1983). The found result in this paper is that red colored teams have an advantage in earning points per game and scoring goals relative to getting goals against. These results can have very important implications for club policy makers who want to change the club colors or for people who want to start up a new football team. According to the results, they should choose the color red as the major color for their home shirts.

# 1. Introduction

Performance is the most important feature in professional sports. Everything goes with performance. For a running athlete the sportive results he achieves will always determine what he will earn and how many sponsors and fans he will attract. For this reason, even minor improvements in the match preparation of an athlete can make major differences in terms of both income and status. In the recent literature, many authors have focused not only on the sportive part of match preparation, but also on the behavioral aspect of the game. This behavioral part is mainly about which colors to wear when participating in a match or contest. The relevant term for this behavioral theme is color psychology. It is how colors influence people's emotions or perceptions and so their resulting behavior. A simple example of the application of color psychology can be found in Glasgow in 2000. In certain areas in this city blue streetlights were installed. Subsequently, the crime rate in these areas decreased. Also a Japanese railroad company followed this move and installed blue lights at one of its stations and therewith it successfully reduced the number of suicide attempts. An explanation for these achievements is that the color blue provokes a feeling of safety and peace in people.

Kaya & Epps (2004) did a research on the relationship between colors and the emotions of people. The main focus was on whether a color was giving people a positive or negative feeling. However, they also asked people to fill in a survey about what colors gave them which emotions. Participants related the color green to the nature and trees. Comfort and soothing emotions were the result together with emotions as happiness, comfort, peace, hope and excitement. Yellow gave participants a lively feeling and energy. Because people often associated yellow with the sun, blooming flowers or the summer time also emotions like happiness and excitement were provoked. Blue was mainly associated with calmness, relaxation, happiness, comfort, peace and hope, but also negative qualified emotions as sadness, loneliness and depression. Red is associated with love and romance, but also with aggression, fight, blood, evil and sometimes even with Satan. White was often related to purity and simplicity. Therefore the color white provoked emotions like innocence, peace and hope, but also emotions as emptiness, loneliness and boredom. Finally, black gave feelings like sadness, depression, fear, anger, death and darkness. Surprisingly, black also brings up the associations richness, wealth and power in people. Additionally, a similar research on the colors purple and grey was done. The relevance of the effects of these colors is minimal regarding the analysis in this paper, so they are not reported here.

An application of color psychology in sports can be found in a paper of Hill & Barton (2005). They randomly assigned blue or red colored uniforms to contestants in several fight sports. The result that they found was that the frequency of winners wearing red was significantly greater than expected by chance. In the same paper they conducted a preliminary analysis on football teams at the European Championships in 2004. They followed five red colored teams and argued that they significantly played better with their red kit compared to playing in a blue or white uniform. In rugby a more extensive analysis was done by Piatti, Savage and Torgler (2012). A positive relationship was found between red shirts and sporting performance in team sport. However, the authors noted that more evidence is needed to draw definitive conclusions on this topic because a lot of confounding factors have to be controlled.

Also an interesting discussion in the world of Judo exists about wearing a blue or white uniform during a match. First, Rowe et. al. (2005) stated that there existed a winning bias at the Olympics of 2004 for judo athletes wearing a blue outfit relative to others who were wearing a white outfit. They suggested that this had to do with the differential effects of the color blue on opponent visibility and/or opponent intimidation. Dijkstra and Preenen (2008) falsified this suggestion and replied with a paper which reported that there was no effect of blue on winning contests in judo. After controlling for confounding factors as allocation biases, asymmetries in prior experience and differences in recovery time they claimed that there exist no differences in blue or white wins.

As can be seen, the discussion about the effect of color on performance is very open. In the world of soccer this is not different. Attrill, Hill & Barton (2008) focused on the long-term success of red-wearing teams in the English Premier League. They found that English football teams with a red home uniform have been champions more often than expected and had a significantly better home record than other colored teams. Kocher and Sutter (2008) conducted an analysis on the performance of different colored teams in the German Bundesliga in the season 2000-2001. They concluded that shirt color could not be the driving factor behind the performance of the German football teams. Compared to individual sports they argued that the effect of shirt color was remarkably lower for team sports like football. Possible explanations for this are the additional effects of team cohesion and/or support by teammates when playing in a team.

This paper will also contribute and focus on the discussion on the effect of team kit color of football teams. As reported above, several analyses have been done in England and Germany, but no general conclusion can be drawn. In the Netherlands, there also exists a professional football league which is called the Eredivisie. This competition is organized since 1956 and through the years 76 teams have been playing in this league. As far as I know, no similar analysis about the Eredivisie is done till now. The question this paper will try to answer is whether different colored uniforms make a difference in the performance of Dutch football teams. The question focuses on the different effects of different colors on performance. The answer on the research question can be very important for new established clubs or for teams who want to change their kit color. If it is the case that a particular color enhances the results of football clubs more than other colors, it would be very interesting for teams to use this particular color in their outfits.

As said, through all the years 76 football teams have played in the Eredivisie. All these 76 clubs will be included in the analysis. Every team has different characteristics and different results. I will use propensity score matching to reveal the treatment effect of a particular uniform color. With the help of the paper of Caliendo & Kopeinig (2008) some practical guidance on implementing propensity score matching was given to me. Because different color categories, i.e. different color kits, exist among football teams, I will conduct different variants of the propensity score matching in which the effects of the different treatments, i.e. different kit colors, compared to the others can be detected.

This paper will follow practically the structure of the *IMRAD* method (**I**ntroduction, **M**ethod, **R**esults **a**nd **D**iscussion). In section 2, I will describe when, where, and how my research was done. For readers who are not interested in the details of the propensity score matching model that is used, section 2.2 can be skipped. In section 3 I will report the results regarding to my research question. In the discussion part I will explain what the implications are and I will end with a small conclusion.

# 2. Method

This part of the paper will go about the used method of investigation. First I will describe how I gathered my data followed by the explanation about the statistical tool I am going to use: propensity score matching.

## 2.1 Data

For the final analysis data on Dutch Eredivisie teams is needed. At www.eredivisiestats.nl a database of all played Eredivisie matches can be found. From the first season (1956-1957) till the most recent season (2013-2014) data on matches played, wins, ties, losses, points, goals made and goals against are available for every team that ever played in the Eredivisie. Thanks to this database I could add a lot of variables to my own dataset containing the home and away record of all clubs ever played in the Eredivisie. The distinction between home and away results is important here since teams usually play in their original kit in home matches and in away matches they often have to switch to an alternative outfit. The home kit does normally not change through the years, while the away outfit changes every year. From the obtained data I have constructed extra variables like points per game and the goals ratio (goals/goals against). These variables will be very important later on in this paper, because they will serve as dependent variables in the statistical analysis. Also control variables as a club's first year in the Eredivisie, if they were the only team of their home city in their first season, if a club was merged or if a club is directly related to other clubs in the dataset. In the next sub-section I will elaborate more on these variables.

Next to the results of all teams, also some other (control) characteristics of the clubs have to be found such as current budget, resident city size and data on home kit colors. Some clubs that do not exist anymore have no data on current budget. This variable will therefore also capture the characteristic whether a football team is abolished or not. If a team is abolished, the current budget variable equals 0. With current budget, budget of the most recent season is meant (season 2013-2014). The size of the budgets of the football teams is always announced at the beginning of each season in several Dutch sport magazines, such as the Dutch magazine Voetbal International. Data on the city of origin for every team can be found at the site of the CBS, which stands for Centraal Bureau voor de Statistiek. This institution keeps track of several statistics about the Netherlands. Data on labor, social security, firms, population, nature etc. is all freely available at their website. With the help of the CBS database the numbers on city size were found. The data on uniform colors is also of crucial matter in investigating the color effects. As known, football teams do not change their home kit colors. They sometimes change the lay-out or change the degree of their basic color in their uniform, but the main color stays always the same. Only when one team merges with another team the color of the outfit changes or can change, but with the merge also a whole new team will be established. The site www.oldfootballshirts.com was the main source for finding the colors of the different teams. On this website, for almost every Dutch team the outfit of past years can be found.

In the following table the number of teams per color is displayed. A distinction is made between main color, second and third color. A team's main color is defined as the major color that is present in a team's outfit. For example, a team's outfit that is for 51% red colored and for 49% white colored, has as major color red and as second color white. Main, second and third color are just determined by the percentage a color is present in a team's outfit. For all teams this distinction was made and that resulted in the following table:

| Color | **Main** | Second | Third |
|---|---|---|---|
| Red | 22 | 1 | 0 |
| Green | 9 | 5 | 0 |
| Yellow | 15 | 5 | 0 |
| Blue | 11 | 3 | 1 |
| White | 12 | 14 | 1 |
| Black | 4 | 4 | 1 |
| Orange | 3 | 0 | 0 |
| *No color* | 0 | 44 | 73 |
| Total | 76 | 76 | 76 |

*Table 1. Number of teams using a particular color in their outfit*

Table 1 describes how many teams wore each color. In the left column all the colors are stated that are present in the dataset. Below in the left column "No color" is stated. In the corresponding row one can see how many teams do not have a main, second or third color. For example, in the row of "No color" there are 44 teams reported at the column "Second". This means 44 teams have no second color. For 73 teams no third color exist.

The analytic tool that will be used in this paper is the propensity score matching method. To make the analysis less complex and easier to conduct, decisions upon which colors can be clustered together in one category have to be made. A second issue that is left is that several teams in the dataset are highly similar. Before having the definitive color categories ready these problems had to be solved.

From the paper of Kaya & Epps, one can derive that the provoked emotions of green and yellow lie close to each other. Lively feelings, happiness and excitement are examples of overlapping emotions of the colors green and yellow. Therefore I decided to put the teams who have green or yellow as their main color in their outfits together in one group. Also the colors blue and white share a lot of similar provoked emotions. Examples are peace and hope. For this reason blue and white will also form a color category. Finally, the two colors black and orange remain, which will be also put in a category under the common denominator 'other'.

After this reallocation of colors into several categories, there also remain some teams that are almost 100% similar to each other. This involves teams that only experienced a name change, while the other characteristics of the team stayed the same. For example, there was a red colored team from Enschede, FC Twente'65, which played its first Eredivisie season in 1965. In 1980, the board of this club decided to remove the '65 from their name. The club went from then on through life under the name FC Twente. Every characteristic or factor of this team stayed the same: color of the outfits, budget, board, trainer, players, etc. It is therefore very likely that the perception people had about this club stayed the same. Clubs that experienced a similar situation were therefore also stacked together with their predecessor as one club. Contrary, teams that, next to a little name change, only had a small change like an additional second color were not 'merged' together with their predecessor. Reason is next to a slightly name change more variables changed. After this reallocation procedure, table 2 on the next page can be constructed.

| Color category | **Major** | Minor |
|---|---|---|
| Red | 18 | 1 |
| Green and Yellow | 19 | 8 |
| Blue and White | 20 | 9 |
| Black and Orange | 7 | 5 |
| *No Color* | 0 | 105 |
| Total | 64 | 128 |

*Table 2. Number of teams per each color category*

We see that three approximately equal color categories are left over, together with a small ‘other’ category consisting of black and orange colored teams. In the final dataset it can be the case that for example blue is the major color of a particular team and white the minor color. In this case this team has no minor color and the major color of this club falls into the blue and white category. The reason why under total 128 minor colors are reported is that also third color is included here. If a club has a second or even a third color, this is often an addition to the most evident color, i.e. the major color of the outfit. Second and third colors are thus qualified as minor colors.

It is also important to remember that clubs do not change their home colors. If a particular football team chooses its base color(s), this color is usually never changed. Only when a merger takes place a team could experience a shirt color change. In this case usually the team’s name also changes. If in case of a merger the shirt color changes, the merged club is considered to be a new team in the dataset. So for instance if Ajax would merge with AZ, the results of the ‘new’ club are included separately from the results of Ajax and AZ.

The final color categories are nice to work with. Namely, with the comparison of these categories there exists a strong link with past literature. The color red is seen to be performance enhancing according to the mentioned paper about rugby of Piatti, Savage and Thaler, but also according to the paper of Attrill, Hill and Barton about the effect of red shirts in the English Premier League. As previously reported, Dijkstra and Preenen argued that there was no effect of blue outfits compared to white outfits on performance in judo. So constructing the blue and white category is also in line with this finding: if there is no significant difference between the two then it makes sense to put them in the same category. Suppose these two colors were significant different, problems could arise since individuals do not interpret them as equal.

## 2.2 Propensity Score Matching

### 2.2.1 Introduction

In the search for a tool to compare the teams in each category I came across the Propensity Score Matching (hereafter: PSM) method. The PSM method was designed in 1983 by Rosenbaum and Rubin. If some researcher wants to find the causal effect of a treatment, he or she usually wants to construct a counterfactual. The PSM method is a tool to find the counterfactual of an observation. In other words, if an observation has outcome $R_0$, PSM can be used to find the outcome that would have occur when a certain treatment was experienced: $R_1$. When estimating the causal effects of a certain treatment there is usually a data missing problem in naturally occurring data since either $R_0$ or $R_1$ is missing. The PSM method constructs either one of the two or both.

In practice the PSM method estimates the following:

- $E(R_1) – E(R_0)$, where E(.) denotes the expectation in the population.

In practice, an untreated individual will be matched with an 'identical' treated individual. The difference in outcome must be due to the treatment then. PSM is especially useful when there is no baseline period available or when subject are not randomly put into treatments as in the case with the football teams colors.

In an ideal situation, an untreated individual perfectly matches a treated individual and it is easy to calculate what the treatment effect is. However, there possibly arise certain problems in pairing up individuals. Bellman (1957) argued that an important issue in matching individuals is the curse of dimensionality. The more dimensions to compare, the harder it becomes to match individuals who are identical in every or almost every aspect. Suppose we have to match people who walk in the city. They will be very different in their characteristics and therefore very hard to find an identical match. Rosenbaum and Rubin show that with their PSM method also this problem can be tackled. The first step in PSM is to estimate the probability that an individual is treated. This can be done by using the probit model. This model was first described by Bliss in 1935. With this probit model a regression can be done with a dependent variable that only can take two values. It can be used to estimate the probability that a certain observation falls into one of the two categories. Then, upon this estimate, individuals will be matched.

Now, a link has to be made with the available data on Eredivisie teams. First of all, one can argue that we have four treatments: shirts that are red colored, green or yellow colored, blue or white colored and other colored shirts. If we can estimate the propensity score of each team in each category, we can match them to another team from a different category to find the counterfactual. In that way we can find the treatment effect, i.e. the effect of a shirt color compared to other colored shirts.

## 2.2.2 The Counterfactual

To successfully construct the counterfactuals, important assumptions have to be made. The *first assumption* of finding the counterfactual is that the treatment group and control group must be the same in absence of the treatment, on average. The assumption is about the homogeneity of the two groups. Unless there seems to be homogeneity here because all the subjects are football teams, this assumption can be very problematic. One can claim that if the dataset goes to infinity the true population average will be revealed, but unobservable factors can cause these different group averages to be different. A simple example can be given using the current dataset on Eredivisie teams. The data consist of red colored teams, but also teams that play in a blue or white outfit. Suppose there is an unobservable factor, and that is whether wealthy people prefer red over blue when they establish a team. As with almost everything in life, money determines the degree of success. If it is the case that there exists an unobservable which makes it more likely that red teams will perform better on average because of more initial wealth, the average of groups can lie far from each other irrespective of the color of each group. So when all the teams had for example no shirts, it would be unlikely that on average the teams in the different categories would be the same. The solution to this problem is to control for these unobservable factors. In the upcoming statistical analysis in this paper control variables are also included. Later on, I will go into more details about controlling. Only, even when I would control for every available variable, there will be always unobservable variables which cannot be controlled. Therefore, while I will control for several confounding effects, I have no illusion that I can perfectly control for every confounding effect and satisfy this assumption completely.

The *second assumption* is that the different groups will respond the same to each treatment in the same way. If we translate this to the application in this paper: all the teams in the different categories should respond the same when they were a particular outfit. FC Utrecht is a famous Dutch team from Utrecht. They play their home matches in red. Another team which often is even as strong as FC Utrecht is the yellow colored team Vitesse from Arnhem. Suppose these teams were matched to each other, then this second assumption states that Vitesse would react the same to a red colored outfit as FC Utrecht does. Assuming that this is the case is probably a bit too simplistic. If the football teams were very homogeneous this assumption could be realistic, but since there are many characteristics for the different teams this assumption seems to be a leap of fate. On the other side, one can argue that football teams are a homogeneous group, because they are equal in lots of views. In that case it would be possible that they would all respond the same to a particular treatment.

The *third assumption* is that the different groups cannot be exposed in isolation to a third factor. Contrary to the first two assumptions, this assumption seems to be more realistic. Think about an economics experiment: the control group only differs from the treatment group because of the treatment. If the two groups are isolated, no other factor could or rather should influence the behavior of the two groups. If a comparison is made between a red team and a blue team and they are matched to each other, it seems unlikely that a third factor, next to the shirt color, will affect one of the two teams. In other words, when two teams are matched to each other, the only thing that differs from them is their shirt color. This third assumption looks like to be equal to the first assumption, but this is clearly not the case. The first assumption states the teams have to be equal in absence of the treatment, while the third one states that if two teams are matched the only factor that distinct the two teams is the color of their shirts and no third factor. Therefore, I consider this third assumption as satisfied.

*The fourth and last assumption* is that unobserved characteristics are equal for treated and untreated. Or in this case: unobserved characteristics are equal for the different treated. As earlier described, unobservable factors are problematic since we cannot assume they are the same for all teams. If we return to the example of wealthier people preferring red over blue when establishing a new football team, it is impossible to assume this is or is not the case. The problem can again be fixed by including as many as possible control variables in the statistical analysis. However, we know that in every econometric model there will always be some error term that consists of unobservable factors. For this reason, the chance that the fourth assumption holds is a twist of fate. I will assume that this assumption is satisfied, but on the other side I recognize that it is possibly been violated.

### 2.2.3 Assumptions of PSM

It would be ideally to match identical teams to each other. However, this cannot be the case since every football team is unique. The PSM method seeks approximate matches as Rosenbaum and Rubin describe in their paper. The PSM method also has several assumptions itself. Let me first discuss these before describing the exact model.

The *first assumption* of the PSM method is the partial equilibrium character, i.e. no general equilibrium effects. This assumption says that the treatment should not indirectly affect the controlled observations. A simple example can be shown by considering thesis workshops given to a limited number of students. Imagine a thesis writing workshop is given to half of a class of students. The ones that received the training are the treated and the ones that did not participate in the workshop are the control group. There might be a risk of spillovers. Students who participated in the thesis workshop could communicate with the other students and pass on their acquired knowledge. Partial equilibrium character does not allow for this and states that the spillover should not influence the outcome of the control group. If we reflect this on the Dutch Eredivisie teams, one can argue that the shirt color of one team does influence another team's outcome. However, the outcome variables in my dataset are the points/game or goals ratio, while a large amount of matches are played by every team. The assignment of a treatment, for example Ajax plays in red shirts, will only have influence other teams' outcome one time per year. Then there are also other factors that influence the outcome of a team, which make the effect of the treatment on the control groups very small and possibly ignorable. The partial equilibrium character is a questionable assumption, however I will assume it is satisfied.

The *second assumption* is the conditional independence assumption. For observational studies, like in this paper, this assumption implies:

- $Y_0, Y_1 \perp D|x$

This equation says that the outcomes $Y_0$, $Y_1$ should be independent of the treatment D conditional on the controlled variables x. So what we need here is that the treatment assignment ignores the outcome. That is, regardless of the performance of teams, the different colors are randomly assigned among teams given their pre-treatment characteristics. In other words, we need the treatment variable to be exogenous. For example, (pre-treatment) good teams should not always wear red colored uniforms and not all (pre-treatment) weaker teams should wear blue colored kits. I think this assumption is satisfied, since one do not know upfront which team will be good or bad. Of course, budget or sources are the main driver behind success, but it is very unlikely that exactly all (pre-treatment) good teams will were a particular color.

The *third assumption* is the matching or overlap assumption. This assumption holds when for each value of x, the control variables or pre-treatment characteristics, there are both treated and non-treated observations:

- $0 < \text{prob}(D = 1|x) < 1$

For each treated observation, there is a matched control observation with similar x. In terms of the current dataset, there has to be found a red colored team with similar characteristics as a blue or white colored team for instance. Although we cannot claim this upfront, it is very likely that this assumption is satisfied given the number of teams in the dataset. In other words, there is a high chance that every club will find a match.

The last and *fourth assumption* is the balancing condition:

- $D \perp x|p(x)$, in which p(x) is the propensity score based on x

The last equation on the previous page claims that the assignment to treatments is independent of the x characteristics, given the same propensity score. Now assume two different colored teams are match, because of the same propensity score. The chance that one team chooses red as major color and the other team choses blue should be independent. Putting this in the context of the Eredivisie teams will mean that when we match two teams based on their pre-treatment characteristics, their choice for a particular colored uniform should be random. Suppose Roda JC, a yellow colored club from Kerkrade, is matched up with AZ, a red colored team from Alkmaar. This mean they have the same propensity score and the choice of their color should be independent. Upfront, one cannot know if this hold. However, there can be tested for it. As discovered later in the statistical analysis section of this paper the balancing condition will hold at all time.

### 2.2.4 Treatment Variables

To set up a model, one needs different kind of variables. The *treatment variables* in this paper are already displayed earlier. The treatment variables are the different colors a team can wear. In table 2, we have seen that four final categories were left after reorganizing the dataset. To recall, this categories were red colored teams, green or yellow colored teams, blue or white colored teams and the 'other' category consisting of black and orange colored teams. For each of these categories, a dummy is included in the dataset. This dummy will indicate whether a club falls into a particular category or not.

### 2.2.5 Control Variables

The *independent* or *control variables* are the current number of residents in home city, the year of first season, a dummy variable which indicates if a team was the only club of the home city in their first season, a dummy variable which indicates of a club is the result of a merger and lastly, a dummy variable which indicates if a club is a direct continuation of another team In the dataset. The control variables are of valuable importance, since the propensity score of the different team will be based on these variables.

First a remark has to be made on possible omitted independent variables. As known, the start budget of a team will be a very important factor in the success of a team. It can for example buy more players and better trainers. Since propensity scores are based upon the control variables, it is kind of a limitation to exclude start budget. It could be the case that clubs with a lot of more resources than other clubs more often choose for red as their base color. There exist three reason why starting budget is excluded in the analysis in this paper. The first one is that there does not exist freely available data on this. The majority of Dutch teams had their first season in the Eredivisie far before the first internet sites were established and could report about budgets or other topics. The information on these budgets is probably somewhere available at the football clubs their private databases, but unfortunately unreachable for me. On top of that, the majority of teams in the dataset do not exist anymore. This makes it even harder to access this information on starting budgets. The second reason why starting budget is excluded is a controversial one. In the past, clubs did often paid out 'dirty money'. These cash flows were not reported to the tax authority in order to pay lower taxes. This was a many performed practice before the professionalization of football in Holland. From my own experience, in the Dutch amateur football competitions it is still a popular way to pay out football players. Because of this way to bypass taxes, the reported starting budgets of teams are possibly biased. A lower budget is probably reported by the tax authority, because a part of the salaries of players is paid out ‘dirty’. The last reason why starting budget is excluded here is because there are good substitutes for this variable. These substitute variables are included in the analysis and are described more extensively on the next pages.

A second omitted variable is an injury variable. This variable indicates when a club has a disappointing season because of injuries. Besides that it is hard to measure this variable, there is probably some parallelism among teams with respect to injuries. The data on Dutch teams go from 1956 till 2014. Because this is a very long time span, I will assume that the number of injuries teams experienced are the same for all teams. This is maybe unrealistic, but because of a long time period I think it is not a problematic assumption.

Next to an injury variable one could also think about an omitted variable like team atmosphere. A trainer can be fired or fights in the squad can occur. Just like an injury variable I will assume that all these possible events are the same for all teams. Therefore a variable on these events is not necessary.

More omitted variables are the quality of the team and the coaching staff, but this variable is already captured by the earlier discussed budget variable.

The first control variable that is used is the current number of residents of home city. It is maybe a bit weird to include this variable because it tells something about the current situation. However, it is an indicator about the fan base of the clubs. It would of course be beneficial to include the number of residents when a club entered the Eredivisie for the first time. Unfortunately, the data on this is not available. The Dutch data institution CBR usually reports about these numbers, but the database on residents only go back to 1988. For this reason the assumption has to made that current resident distribution is the same as when a club entered the Eredivisie for the first time. The distribution is not exactly the same, but it is not very likely that one city that was one of the smallest in 50s is now one of the largest. Only the differences between cities are probably smaller or lower than formerly. With this control variable the effects of fan base and resources are captured. One would expect that when the number of residents in the home city is high for a club, there is a higher potential for supporters for this club. If there are more fans, the attendance when a team plays home will be higher and therefore their home advantage. This can also be due to a bigger stadium, since there is a possibility to house more supporters. When the number of residents is very high, there are probably also more resources. There is namely a higher chance for sponsors, but also higher revenues from ticket sales and merchandising. Therefore the control variable current number of residents of home city will also take over the function of the omitted variable starting budget.

The second control variable is when the first season is played by a particular team. In 1956 the Eredivisie started, so when a team entered the Eredivisie in that year the first season variable equals 1. When a team entered the Eredivisie for in the third season (1958), this variable equals 3. This variable measures timing effects such as fashion, economy and experience effects. It can be that in a certain year it was a trend to wear blue. Suppose this was the year 1965, then it could be the case that the color of teams that were founded in 1965 are more times blue than another color. Also economy effects are considered with this variable. Assuming that in the 70s the economy in Holland was booming, teams that are established in this time period have probably a higher starting budget then time that are established in another time period. Experience is also a factor that is measured with this first season variable, because the later a team is established the lower their experience today. The relationship between performance and year of first season is a bit vague upfront. The direction could go both ways.

The third control variable is a dummy variable which indicates if a team was the only club of their home city in their first season. If a club is the only team in their first season, they have probably a higher budget compared to being not the only team. Two clubs from the same city in one competition have to share the fan base and resources in the city. For this reason, a club that enters the Eredivisie as the only club of their city has more resources at its disposal. Therefore an expected positive relationship between performance and this control variable exists. Thereby, this variable will also partly take over the effect of a budget variable.

The fourth control variable is a dummy variable which indicates of a club is the result of a merger. This variable will measure the effect of a club being the result of a merger between two clubs. A merger results in an increased quality of the squad of a club plus the fan base and resources will increase as well. Two clubs will create one squad, so only the good players of both clubs will remain. Performance is likely to be better then. Also the fan base will be bigger since the clubs are now forming one club. This also holds for the resources which will become more. The number of sponsors will increase since two groups of sponsors are now pooled together. This merger variable is expected to have a positive effect on performance and also takes over part of the effect of a possible budget variable.

The fifth and last control variable is a dummy variable which indicates if a club is a direct continuation of another team in the dataset. In the dataset reorganization two clubs were clustered together when one was an exact continuation of the other. However, there were also teams that were an exact continuation but some variable(s) changed. GVAV is a Dutch team from Groningen and its name was changed in FC Groningen in 1961. Next to a name change, there was also a change in uniform color. From a completely green shirt, the outfit changed to a white shirt with some green stripes. For these kind of teams this direct continuation variable is designed. A direct continuation points to the fact that a team knows already the Eredivisie, so has more experience, but also that they have the resources to play in the Eredivisie.

### 2.2.6 Dependent Variables

The dependent variables can also be defined as the performance variables. The control variables and the treatment variables will determine these performance variables. In this paper I will use the two dependent variables that are described below.

The first performance variable is points per game. All the different teams have not played the same amount of matches. Ajax from Amsterdam, Feyenoord from Rotterdam and PSV from Eindhoven have all played the most matches in the Eredivisie. This is because they have played continuously in the Eredivisie from the establishment of this competition. But because of this fact, they also scored the most points in the Eredivisie. Since other teams are also included in the analysis that have played a lot fewer matches, a relative measure is needed. The most simple variable to construct is the variable points per game. In football, a team earns three points when it wins, one point when it ties and zero points when it loses. In the descriptive analysis later on, numbers on this variable will be showed.

The second performance variable is the goals ratio. A team can earn points for a match, but in order to do this they need to score goals and have the least possible number of goals against. Goals and goals against can therefore also be seen as performance variables. By putting these two measures together I constructed the goals ratio. It measures how much goals a team scores per goal against. The ratio is simply defined by dividing the goals scored by the goals against. Again the three clubs that had the most games played had the most goals scored and the most goals against. In order to compare them with other teams in terms of goals, the goals ratio is constructed.

## 2.2.7 Propensity Scores

The general model of the PSM method exists of three parts. First the propensity scores have to be calculated. Secondly, teams have to be matched that have the same propensity score. The last step is the calculation of the treatment effect using one particular group and their matched counterfactuals.

As earlier said, first a probit model will be estimated. The probability of an observation having the treatment will be calculated with this model. In general form this looks like:

- $S(D = 1) = \beta x'$

The term S(.) stands for the propensity score, D is the treatment variable and β is a vector of regression coefficients which measure the effects of the control variables vector x'. Translated to the data in this paper, there will be different propensity scores for the teams. First a propensity score when the shirt color is red, one for green or yellow colored shirts, then one for a blue or white colored shirt, and lastly one for black or orange shirts. In mathematical terms this boils down to:

- $S(D = \text{Red}) = \beta_{R,0} + \beta_{R,1}\text{CurrentResident} + \beta_{R,2}\text{FirstSeason} + \beta_{R,3}\text{OnlyFromCity} + \beta_{R,4}\text{Merged} + \beta_{R,5}\text{Direct}$
- $S(D = \text{Green or Yellow}) = \beta_{GY,0} + \beta_{GY,1}\text{CurrentResident} + \beta_{GY,2}\text{FirstSeason} + \beta_{GY,3}\text{OnlyFromCity} + \beta_{GY,4}\text{Merged} + \beta_{GY,5}\text{Direct}$
- $S(D = \text{Blue or White}) = \beta_{BW,0} + \beta_{BW,1}\text{CurrentResident} + \beta_{BW,2}\text{FirstSeason} + \beta_{BW,3}\text{OnlyFromCity} + \beta_{BW,4}\text{Merged} + \beta_{BW,5}\text{Direct}$
- $S(D = \text{Black or Orange}) = \beta_{O,0} + \beta_{O,1}\text{CurrentResident} + \beta_{O,2}\text{FirstSeason} + \beta_{O,3}\text{OnlyFromCity} + \beta_{O,4}\text{Merged} + \beta_{O,5}\text{Direct}$

We see that the probability of a team having a particular shirt color is determined by the earlier explained control variables current residents of city, year of first season, only team from city in first season, merged and direct continuation. For all the teams in the dataset these four scores were calculated: S(Red), S(Green or Yellow), S(Blue or White) and S(Black or Orange). If we want to calculate the treatment effect of a red colored outfit, we have to match up red colored teams with other colored team that have the same propensity score. There are multiple ways to match the different teams. In this paper I will discuss and use four methods.

## 2.2.8 Types of Matching

The first one is nearest neighbor matching. In this method, the absolute differences between the estimated propensity scores for the control and treatment groups is minimized. The control and treatment subjects are randomly ordered. Then the first treated subject is selected along with a control subject with a propensity score closest in value to it:

- $C(P_i) = \min|P_i - P_j|$

$C(P_i)$ represents the group of control subjects *j* matched to treated subjects *I* (on the estimated propensity score). $P_i$ is the propensity score of treated subject *i* and $P_j$ is the propensity score of control subject *j*. So for example, a red colored team will be match to a different colored team which propensity score is closest to the propensity score of that red colored team. In simple words, the nearest neighbor of a particular colored club will be found and matched to this other colored neighbor. This method lies the closest to perfect matching. With perfect matching, two teams are match that have the exact same propensity score. However, in practice this is hardly the case. Nearest neighbor matching the best alternative since it will match up two teams that have propensity scores as closest as possible to each other.

The second variant of propensity score matching is kernel matching. In this method, every treated subject is matched with the weighted average of the control subjects. The weights are inversely proportional to the distance between the treated and control group's propensity scores. So a red colored team will be matched to the weighted average of the other teams. This weighted average is based on the propensity scores of these other colored teams. If the propensity score of other colored team lies far from the propensity score that red colored team, the average of the other colored team gets less weight. Kernel matching can be displayed as follows:

- $w(i,j) = \frac{K(\frac{Pj-Pi}{h})}{\sum_{J=1}^{no} K(\frac{Pj-Pi}{h})}$

Here h is the bandwidth parameter. A better example of this method can be given. Suppose we want to match up RKC Waalwijk, a yellow colored team from the south of Holland. The propensity scores of all other teams are considered, and based on the distance to the propensity score of RKC Waalwijk, a weight is calculated. The farther a propensity score an another team lies from the propensity score of RKC Waalwijk, the less weight it gets. Then the outcome (or performance) of the other teams are multiplied by this weight and compared to the outcome of RKC Waalwijk. In this way a possible treatment effect can be discovered. The advantage of including this method in this paper is that it not only considers the outcome of one matched counterpart of RKC Waalwijk, but considers the outcomes of all teams.

The third method of propensity score matching is radius matching. In this method, every treated subject is matched with a corresponding control subject that is within a predefined interval of the treatment subject's propensity score. Since each of the treatment subjects must be matched with a control subject within a given interval, only a certain number of comparisons will be available. For instance, a red colored team will be match with a different colored team within a radius of 0,1. This means the propensity score of a matched counterpart of a red colored team have to lie within 0,1 probability of the propensity score of the red colored team. The radius matching method can be described by the following:

- $|| P_i - P_j || < r$, where r is the radius parameter.

The reason to include this type of probability matching is that for comparing the outcome of one club with other teams, only the others teams are considered that have the closest propensity score to that one club. It depends on the value of the radius parameter how many other teams are used to compare outcomes with.

The last method is stratification matching. In this method the outcomes will be compared within intervals/blocks of propensity scores. The propensity scores are classified into intervals based on the range of values. Each interval consists of treatment and control subjects that on average, have equivalent propensity scores. The differences between the outcomes of the treatment and the control group are calculated to obtain the average treatment effect. It is an average of the outcomes of a treatment per block weighted by the distribution of treated subjects across the blocks. This method is a bit an unusual type of matching and not widely used. It is included in this paper to see whether it can answer the question if there exist different effects of shirt color on the performance of different teams.

### 2.2.9 Treatment Effects

Now the several matching methods are described, we can start with measuring the treatment effects. After matching, there exist two groups: the treated group and their counterfactuals. For example, the red colored teams and there counterfactuals. Then the calculation for the average treatment effect of a red kit can be conducted:

- E(S(Red)|D = Red) – E(S(Red)|D = No Red) = Average treatment effect of color red

This can of course also be done with the teams in the other color categories:

- E(S(Green or Yellow)|D = Green or Yellow) – E(S(Green or Yellow)|D = No Green or Yellow) = Average treatment effect of color Green and Yellow
- E(S(Blue or White)|D = Blue or White) – E(S(Blue or White)|D = No Blue or White) = Average treatment effect of color Blue or White
- E(S(Black or Orange)|D = Black or Orange) – E(S(Black or Orange)|D = No Black or Orange) = Average treatment effect of color Black and Orange

E(.) is the outcome or level of the performance variable. In the statistical analysis all these four average treatment effects will be estimated. There exist two kinds of results within a performance variable: one result in home matches and one in away matches. Suppose there exist a significant treatment effect for a red colored outfit using the home record, one should also control for the away record. In the away matches, the teams usually do not play in their home kits. If the treatment effect for a red colored outfit is then also present for the away record, we can argue that the significant treatment effect of the red color in the home record is due to omitted variables. In a later section, I will check if there exist a significant treatment effect when using the home and away record.

## 2.3 Motivation for using PSM

From the earlier sections one can learn that PSM is the right method for analysis in this paper. The first reason is that PSM is proven to be a successful tool to reveal the effects of a particular treatment. PSM is also widely used when the subjects are not randomly put into treatments. This is the case with the football teams in the dataset. The teams were not randomly assigned to a particular shirt color, but they have chosen their team colors themselves. In subsection 2.2.3, we also saw that the practice in this paper is not violating the PSM assumptions. This is very important since it proves PSM can be used in this paper. Next to this, the dataset also consist of enough control variables for accurately calculate the propensity score of all football teams. Therefore the PSM method is suitable for being the method of analysis in this paper.

# 3. Results

## 3.1 Descriptive Analysis

In the descriptive analysis I will provide some preliminary statistics. The means for the continuous variable and the frequency for the dummy variables will be reported.

In the previous section four color categories could distinguished. For all of these categories some descriptive statistics are described in the following table:

| ***Color Category :*** | **Red** | **Green and Yellow** | **Blue and White** | **Black and Orange** | **Total** |
|---|---|---|---|---|---|
| Total # | 18 | 19 | 20 | 7 | 64 |
| Mean # of Residents | 259130 | 177375 | 363112 | 155795 | 256051 |
| Mean First Season | 8,1 | 18,8 | 12,0 | 19,1 | 13,7 |
| # Only from City | 14 | 15 | 10 | 6 | 64 |
| # Merged | 3 | 6 | 2 | 0 | 64 |
| # Direct | 1 | 4 | 3 | 1 | 64 |
| Mean Points/Game | | | | | |
| *Home* | 1,68 | 1,38 | 1,43 | 1,33 | 1,47 |
| *Away* | 1,00 | 0,76 | 0,84 | 0,83 | 0,86 |
| Mean Goals Ratio | | | | | |
| *Home* | 1,57 | 1,06 | 1,08 | 1,02 | 1,20 |
| Away | 0,75 | 0,49 | 0,54 | 0,54 | 0,58 |

*Table 3. Descriptive statistics of all variables*

From table 3, one can see the different variables in the left column. The color categories are in the top row. For every variable the highest number is underlined. The first variable is the number of teams per category. We can see that there are three categories of approximately the same size. The most teams wear the color blue or white. The number of teams in the black and orange category is clearly much lower than the number of teams in the other categories.

### 3.1.1 Mean # of Residents

In the second row we have the mean number of residents in the home city. In the blue and white category this mean is the highest. This suggests that the blue or white colored teams have a higher fan base and possibly more resources. The mean number of residents of the red colored team lies slightly above the average, while for the other two categories the mean number of residents lies much lower.

### 3.1.2 Mean First Season

The mean of year in the first season is in the third row. This variable indicates in which season a team had it first season in the Eredivisie with 1956 as the base year. The teams in the black and orange category have their first season in the Eredivisie the latest on average, closely followed by the green and yellow colored teams. The red colored teams have on average their first season as first, which suggest that the color of red was more popular in the earlier years of the Eredivisie. The number 8,1 means that on average the first season of red colored teams was in 1963 (1956 = 1, 1957 = 2, etc).

### 3.1.3 # Only from City

The fourth row displays the number of times a club was the only team in the city when entering the Eredivisie for the first time. The green and yellow colored teams experienced this situation the most, followed by the red colored teams. When a team enters the Eredivisie and it is the only club from the home city it suggests it has more resources and fan base available since it does not need to 'share the city' with another team.

### 3.1.4 # Merged

The fifth row indicates the number of merged teams per category. From the numbers in this row it can be seen that green and yellow colored teams were merged twice as much as red colored teams, while the red colored category is ranked secondly with respect to mergers. If a team is merged, it has usual more resources, a larger fan base and a better squad since it is basically two teams in one club.

### 3.1.5 # Direct

The next variable listed in table 3 is the number of clubs being a direct continuation of another team. Again, the green and yellow colored teams are most often a direct continuation of another team. In the dataset there are 3 blue or white colored teams that are a direct continuation of another team in the dataset.

### 3.1.6 Mean Points/Game & Mean Goals Ratio

Then the numbers in the last four rows of table 3 are averages of the performance variables, or dependent variables. One can observe that red colored teams are scoring quite well on these variables. This holds for both home and away record. After the teams in the red category, the teams in the blue or white category perform the best. They lie on average far behind the teams in the red category, but they score on every performance variable better than teams in the green or yellow category and teams in the black or orange category. For the teams in these last two categories, It is not straightforward to say which teams perform better. What we can say is that all the performance measures of these two last categories lie below the total average, both for home and away record. From these numbers a trend of red colored teams performing can be discovered. However, before a conclusion can be drawn a more extensive analysis is needed.

## 3.2 Statistical Analysis

In this statistical part, I am going to conduct the PSM method with the available data. First a PSM model with the points per game as the performance variable will be estimated for the home and away record. Thereafter the same will be done, but then the goals ratio will serve as the performance variable. In this paper, STATA is used to estimate the several propensity score matching models.

### 3.2.1 Propensity Score Matching with Points per Game

Now I will start estimating several models with points per game as the dependent variable. Before I present the test results, I will first give some pre-explanation of some features of the analysis. For example, I used common support for matching the different colored teams. Common support means that when there are two groups, the treatment group and the control group, there exist a range of propensity scores for each group. For instance the lowest propensity score of the treatment group is 0,1 and the highest score is 0,9. If the range of propensity scores of the control lies in the interval [0,15;0,95], the common support region lies in the interval [0,15;0,9]. The reason to use common support is that there exist teams in either the treatment or control group that do not have a counterfactual in the other group. In a way, one could consider these teams as outliers. Thereby, using common support is normally only seen as a problem when the observations of the treatment and control group are very different. Since this problem does not exist here, I will use common support for matching the several teams. A second, unusual feature that is used in the analysis is the use of bootstrapped standard errors when matching teams with the kernel matching method. Bootstrapped standard errors can be used when the theoretical calculation of the standard errors fails like in the case of kernel matching. With bootstrapping, random samples of the sample dataset are drawn repeatedly. Based on these random samples one can calculate the sample standard deviation of the sampling distribution. Because only the standard errors with kernel matching cannot be calculated analytically, I will only use bootstrapped standard errors when conducting this type of matching.

#### 3.2.1.1 Red – Points per Game

The first treatment category that is analyzed is the category with the red colored teams. Red teams are matched to a different colored team to find the effect of a red colored uniform on the points per game. In the next table, the results can be found. This table 4 shows six different models: a t-test, a t-test with control variables and the four matching methods. The dependent variable is points per game in all models.

When a t-test is conducted, the average treatment effect of a red colored outfit is 0,28. This effect is significant on a 1% significance level. This means that when a team has a red colored shirt, their points per game increase by 0,28. This t-test can be viewed as a simple regression with only a dummy variable included on whether a team is red colored or not. When comparing the points per game in the home matches with the away matches, the color red is of less importance in away matches. This is because the effect of are red colored shirt on points per game in away matches is only 0,20, which is significant on a 5% level. However, there are reasons why we cannot say that the effect is higher in home or away matches with this t-test. In away matches the teams usually do not wear their home colors. So a comparison between the effects in home and away matches can be made, but it makes

| **Red - Points per Game** | Average Treatment effect | |
|---|---|---|
| Model | *Home* | *Away* |
| T-test | 0,28*** | 0,20** |
| T-test with control variables | 0,23** | 0,17* |
| Nearest Neighbor Matching | 0,17 | 0,12 |
| Kernel Matching | 0,21** | 0,15 |
| Radius Matching | 0,22** | 0,15 |
| Stratification Matching | 0,19* | 0,13 |

*** = Significant on a 1% level

** = Significant on a 5% level

* = Significant on a 10% level

*Table 4. The average treatment effects of red colored shirts on points per game*

no sense to say that the effect of a red colored shirt is higher or lower in away matches than in home matches. The treatment effect in away matches is only reported such that it is observable if the treatment effect in home matches is not due to unobservable factors. It is only straightforward that when the treatment effect in both home and away matches is equal in terms of significance, one can argue that it is not the treatment that drives the results but rather unobservable variables. However, because no control variables are included in the t-test, no firm conclusions can be made about the effect of red colored shirts. If inferences have to be made, we need to control for other variables that can drive performance. Therefore the t-test is only included in this table for strictly informative purposes.

The next model that is estimated is a t-test while controlling for the control variables earlier described in this paper. In this controlled t-test the effect of a red colored kit in the home match is 0,23 which is significant on a 5% level. The points per game of a red colored team will be 0,23 higher than the points per game for a team with another color according to this result. A look at the average treatment effect in away matches reveals that the effect of a red kit in these away matches are 0,17 higher compared to other colored teams. This effect cannot be taken seriously, since teams do not wear their home colors in away matches normally. Still, if control variables are included in a regression, there exists a treatment effect in away matches. This is probably due to omitted unobservable variables. Hence, a t-test with the controlled variable will also be considered as informative and so causal inferences cannot be made based on this model. If conclusions about the treatment effects have to be made, the average treatment effects of the matching methods have to be considered.

In the nearest neighbor matching method one can see that the average treatment effect of the color red is 0,17. However, this effect is not even significant on a 10% level. If only nearest neighbor matching was considered, a possible treatment effect of a red colored outfit should be rejected. The average treatment effect is also insignificant for the away record of the red teams. When looking at the kernel matching method with bootstrapped standard errors, the average effect of a red colored shirt is significant on a 5% level. This suggest that red shirts ensure for 0,21 points per game more than teams that do not were red shirts. A quick look at the treatment effect in away matches, it can be seen that there is no significant effect of shirt color. This is what we would expect if we believed that there exist a significant effect of wearing red shirts. For radius matching and stratification matching one could observe the same: a significant effect of red shirts in their home matches and no effect in away matches. This totally makes sense if it is true that red shirts enhance performance, because in away matches the teams usually do not wear their base colors. According to the radius matching method, wearing red shirts in home matches results in a 0,22 significant increase in points per game compared to wearing other colored shirts (significant on a 5% significance level). When using the stratification method, this significant effect equals 0,19 on a 10% significance level.

### 3.2.1.2 Green and Yellow – Points per Game

For the next category the same can be done. This next category consists of green and yellow colored teams. Below the several models are showed in a table with points per game as the dependent performance variable.

| **Green and Yellow - Points per Game** | Average Treatment effect | |
|---|---|---|
| Model | *Home* | *Away* |
| T-test | -0,14 | -0,15* |
| T-test with control variables | -0,10 | -0,10 |
| Nearest Neighbor Matching | -0,27** | -0,24** |
| Kernel Matching | -0,16 | -0,15* |
| Radius Matching | -0,06 | -0,08 |
| Stratification Matching | -0,18** | -0,16** |

*** = Significant on a 1% level

** = Significant on a 5% level

* = Significant on a 10% level

*Table 5. The average treatment effects of green or yellow colored shirts on points per game*

In table 5 the effects of green or yellow colored shirts on point per game are displayed for the six estimated models. The t-test, a simple regression model with one dummy variable for the treatment, gives a controversial result. In the home matches there is no significant treatment effect for the green and yellow colored teams, but in the away matches there is. However, the t-test is just a preliminary informative result, and the effect of green or yellow shirts is strange to exist in away matches. It suggests that green and yellow teams are performing significantly better than other colored teams in away matches. This effect has to be due to not included independent variables, since green and yellow colored clubs do not wear these colors in their away matches. When including the control variables in the next model, the treatment effects are insignificant in both home and away matches.

When going to the different PSM methods, using the nearest neighbor matching method results in an average treatment effect of -0,27 (significant on a 5% level). This suggests that when a team wears a green or yellow outfit, their points per game are decreased by 0,27 compared to other colored teams. This result indicates that green or yellow colored teams have a disadvantage compared to other teams in terms of shirt color. However, when checking if the same effect exists when considering the away match record, the answer is yes. The average treatment effect is -0,24 and also significant on a 5% level. The significant effect of a green or yellow colored shirt is therefore probably due to unobservable not included factors instead of color. For the kernel and radius matching methods, there does not exist a significant effect of the green and yellow shirts. With kernel matching there exist however a significant average treatment effect on points per game when considering the away matches. Since the away matches are usually not played in the home colors, I assume that this effect is due to other (omitted) factors. Finally, with stratification matching, there also exists a significant negative average treatment effect of a green or yellow colored shirt, but since the effect in away matches is also significant, this effect cannot be assigned to the shirt color.

### 3.2.1.3 Blue and White – Points per Game

The third color category consists of blue and white colored teams. Also the treatment effect of these colors are tried to be estimated using the six models. Below a table is displayed with the results.

| **Blue and White - Points per Game** | Average Treatment effect | |
|---|---|---|
| Model | *Home* | *Away* |
| T-test | -0,06 | -0,03 |
| T-test with control variables | -0,08 | -0,07 |
| Nearest Neighbor Matching | -0,17 | -0,21 |
| Kernel Matching | -0,10 | -0,10 |
| Radius Matching | 0,04 | 0,04 |
| Stratification Matching | -0,11 | -0,09 |

*** = Significant on a 1% level

** = Significant on a 5% level

* = Significant on a 10% level

*Table 6. The average treatment effects of blue and white colored shirts on points per game*

Above the table of the average treatments effects of blue and white colored shirts on points per game is stated. About the results a short conclusion can be made: in none of the models the average treatment effect of a blue of white shirt is significant. Even in the t-test without any control variables there is no significant effect. When putting the control variables into the regression together with the treatment dummy, the effect on points per game stays insignificant. For all matching methods, there is no significant treatment effect of blue or white colored shirts. This holds for both home and away matches. The results of these models mean that there probably does not exist an advantage or disadvantage of wearing blue or white colored uniforms when playing matches.

### 3.2.1.4 Black and Orange – Points per Game

The fourth category that is analyzed is the black and orange category. The effects of black and orange shirts are put into table 7.

| **Black and Orange - Points per Game** | Average Treatment effect | |
|---|---|---|
| Model | *Home* | *Away* |
| T-test | -0,16 | -0,04 |
| T-test with control variables | -0,08 | -0,03 |
| Nearest Neighbor Matching | -0,07 | 0,04 |
| Kernel Matching | -0,07 | 0,05 |
| Radius Matching | -0,07 | 0,06 |
| Stratification Matching | -0,06 | 0,06 |

*** = Significant on a 1% level

** = Significant on a 5% level

* = Significant on a 10% level

*Table 7. The average treatment effects of black and orange colored shirts on points per game*

Again as with the blue and white colored teams, there seems to be no effect of the shirt colors black and orange on points per game. Shortly: there is nog significant effect found in all models for both home and away matches. For away matches this is expected, but if there would be a color advantage or disadvantage of black and orange shirts, the average treatment effect in home matches should be significant. This is not the case, so neither an advantage nor a disadvantage on points per game exists for black and orange team when considering the six reported models.

## 3.2.2 Propensity Score Matching with Goals Ratio

The second performance variable, or dependent variable, is the goals ratio. The ratio is equal to goals made divided by the number of goal against. The effects of the different colors on this performance variable are estimated.

### 3.2.2.1 Red – Goals Ratio

| **Red - Goals Ratio** | Average Treatment effect | |
|---|---|---|
| Model | *Home* | *Away* |
| T-test | 0,51*** | 0,23*** |
| T-test with control variables | 0,46** | 0,21*** |
| Nearest Neighbor Matching | 0,43** | 0,17* |
| Kernel Matching | 0,44** | 0,19** |
| Radius Matching | 0,42** | 0,19* |
| Stratification Matching | 0,40* | 0,17* |

*** = Significant on a 1% level

** = Significant on a 5% level

* = Significant on a 10% level

*Table 8. The average treatment effects of red colored shirts on the goals ratio*

In table 8, the six models are reported together with their estimate of the effect of red colored outfits on the goals ratio. In the simple t-test one can observe a significant treatment effect of 0,51. This means that red colored teams score per goals against 0,51 goals more than other colored teams. However, in this t-test no control variables are included, so the effect can also be driven by omitted variables. On top of that there is also a significant treatment effect on away matches, which strongly suggest that the goals ratio is driven by something else than the treatment variable. When controlling using the control variables earlier described in this paper, the treatment effect stays significant on a 5% level. On the other side this also holds for the away matches, so no causal inferences about the color red can be made here. If we continue to the four matching methods, we see the average treatment effect in the nearest neighbor method is significant on a 5% level. Its size is 0,43 which means a red wearing team scores 0,43 goals more per goals against relative to other colored teams. In contract to this, the significant effect of red shirts is also present in the away matches. It is not a strong as the effect in home matches because it is only significant on a 10% level, but it suggests that other factors could also be the driver of the goals ratio of red teams. In this case there cannot be an exclusion of the fact that red shirts drive performance on the goals ratio, since the treatment variable in away matches is of less explanatory power. For radius matching the same holds. In home matches the red colored shirts have a significant effect, but in away matches also a significant treatment effect exists although this effect is only significant on a 10% level. If one argues that significance on a 10% level does not prove strong explanatory power, one could state that the effect of red colored shirts is certainly present. For kernel and stratification matching in both home and away matches there exist significant treatment effects on the same significance level. As we know, in away matches the home kit is normally not worn by the team so a significant effect should not exist in away matches if one wants to prove an advantage of wearing red shirts.

### 3.2.2.2 Green and Yellow – Goals Ratio

The second treatment variable is a dummy about whether teams wear green or yellow colored shirts. The effects of green and yellow colored shirts are also estimated.

| **Green and Yellow - Goals Ratio** | Average Treatment effect | |
|---|---|---|
| Model | *Home* | *Away* |
| T-test | -0,20 | -0,14** |
| T-test with control variables | -0,16 | -0,09 |
| Nearest Neighbor Matching | -0,38** | -0,18** |
| Kernel Matching | -0,26* | -0,13** |
| Radius Matching | -0,08 | -0,08 |
| Stratification Matching | -0,26** | -0,13** |

*** = Significant on a 1% level

** = Significant on a 5% level

* = Significant on a 10% level

*Table 9. The average treatment effects of green and yellow colored shirts on the goals ratio*

In table 9 the results are stated. For the t-test without the control variables no significant effect is found in home matches for wearing green or yellow shirts. The weird thing is that in away matches a significant treatment effect is present. If the control variables are put in the model, the significant treatment effect in away matches also disappears. A further look to the nearest neighbor matching method, the effect of green and yellow colored shirts is significant on a 5% level and equals -0,38. This would mean that teams that wear green or yellow shirts score 0,38 goals less per goal against than teams that were another color. This suggests a strong disadvantage compared to other colored teams. However, the significant treatment effect holds also for the away matches on the same significance level (5%). If no other factors but shirt color would drive the goals ratio, it would be normal to have no or a less significant treatment effect for the played away matches. For kernel and stratification matching holds the same situation. A significant treatment effect exists in the home matches, but it does not disappear when considering the away matches. This suggests that the disadvantage on the goals ratio for green or yellow colored team is driven by another factor. When considering radius matching, no significant effect in both home and away matches is found. Therefore no statements can be made about an advantage or disadvantage of wearing green or yellow in football matches.

### 3.2.2.3 Blue and White – Goals Ratio

Next, the effects of blue and white shirts on the goal ratio are examined. On top of the next page a table can be found with the results using six different models.

In table 10 the results are reported. When a regression is run with the dummy variable of whether a team wears green of yellow as only dependent variable, there exists no significant average treatment

| **Blue and White - Goals Ratio** | Average Treatment effect | |
|---|---|---|
| Model | *Home* | *Away* |
| T-test | -0,19 | -0,06 |
| T-test with control variables | -0,24 | -0,11 |
| Nearest Neighbor Matching | -0,61** | -0,31** |
| Kernel Matching | -0,34 | -0,15 |
| Radius Matching | -0,06 | -0,02 |
| Stratification Matching | -0,32 | -0,14 |

*** = Significant on a 1% level

** = Significant on a 5% level

* = Significant on a 10% level

*Table 10. The average treatment effects of blue and white colored shirts on the goals ratio*

effect in home matches. In away matches this treatment effect is also insignificant, but this makes perfectly sense since the home kits are usually not worn in the away matches. When putting the control variables into the regression model, the results remain the same. Considering the different matching methods, only a significant average treatment effect can be discovered when using nearest neighbor matching. The size of this negative effect of wearing blue or white is pretty large. Using blue or white as a shirt color would decrease per goals against score goals by 0,61. However, the significant effect is also present in the away matches. Therefore, one can assume that this significant effect in home matches is not driven by shirt color, but by other third factors. In the three other matching methods no significant treatment effect was found for both home and away matches.

### 3.2.2.4 Black and Orange – Goals Ratio

The effect of the last color category black and orange on the goals ratio is also evaluated. The following table reports the results:

| **Black and Orange - Goals Ratio** | Average Treatment effect | |
|---|---|---|
| Model | *Home* | *Away* |
| T-test | -0,21 | 0,05 |
| T-test with control variables | -0,10 | 0,01 |
| Nearest Neighbor Matching | -0,10 | 0,01 |
| Kernel Matching | -0,05 | 0,01 |
| Radius Matching | -0,03 | 0,02 |
| Stratification Matching | -0,03 | 0,02 |

*** = Significant on a 1% level

** = Significant on a 5% level

* = Significant on a 10% level

*Table 11. The average treatment effects of black and orange colored shirts on the goals ratio*

Again, for black and orange shirts in none of the models there exists a significant average treatment effect on the performance variable. This was also the case when evaluating the effects on points per game and now it is also the situation with the effect on the goals ratio. This is probably partly due to a very low number of observations. The number of teams in this category is namely seven. According to the results no causal inferences can be made, since they are all insignificant. If one would expect a significant effect of black or orange shirts an insignificant effect of black and orange shirts in away matches would be evidence. However, there does not exist a significant average treatment effect in none of the six models so a color effect of black and orange seems to be not present.

# 4. Discussion

According to the previous section, important implications on both the results and the different models can be made. They will be discussed in this section together with some remarks on limitations and future research.

## 4.1 Implications

From the results section one can learn that the effects of the color category were different for the two different performance variables. Regarding the points per game only for the red and green or yellow colored teams an effect of their shirt color was found. However, for the green and yellow teams this also held for the away matches. In these matches usually the away kit, which differs from the home kit, is used. For red colored team the effect of the shirt color disappeared when the away matches were considered. Therefore, red colored teams have an important advantage in gaining more points per game. This would mean that it would be very beneficial for teams to use red colored shirts, since wearing these shirts yields a benefit compared to other teams. When a team is gaining more points per game it is more likely that they will win prizes. Becoming champion of the Eredivisie or winning a cup will become easier only by wearing red shirts. The costs of changing to a red outfit are mainly historical costs. The color change would not be cost rising in terms of production costs, but it could be hard for teams to change their base colors which they have worn for a long time. The identity of the club will be affected. Sponsors, fans and all people who know the club associate the team with their home color. A change in color gives new meaning to a club, but this could also turn out good. Looking to the effects of wearing red shirts, it is at least 0,19 and at most 0,22 points per game according to all matching methods except the nearest neighbor matching method. A team plays per season seventeen times in their home outfit plus a few times in the away matches. Suppose some team plays in a red colored outfit twenty times per season. This would mean they will earn 3,8 to 4,4 points more on average than other teams in one season. The teams that do not wear red shirts have to make the tradeoff between the increase in points and the costs that are incurred with a color change. For teams that are new established it is certainly a good idea to use red as their shirt color. If they would use another color the cost would be the same, but the performance in terms of points per game would probably turn out lower.

For the goals ratio it seems to be the case that in away matches a systematically effect, advantage or disadvantage, is present in all models for red and blue or white colored teams. This is a weird fact since no team uses its home kit in every away game. When considering the goals ratio, the effects of green or yellow and black or orange colored shirts are negligible. For the blue or white team the present effect of shirt color in away matches makes the present treatment effect in home matches unreliable. For red colored teams for every matching method, except for stratification and kernel matching, the color effect in away matches is of less explanatory power than in home matches. If significance on a 10% level would be considered as insignificant, the effect of red colored shirts proves its value. In this case approximately 0,425 goals per goal against are made on average more by red colored teams compared to other colored teams. A given fact is that if a team scores more goals than its opponent it will win the game. Other than red colored teams with a mean goals ratio of above 0,575 should have serious interest in changing their colors to red. If these teams were red they will on average score more goals than their opponent (0,425 + goals ratio above 0,575 > 1). By winning more matches, more success will be the result. Also generating more money because of more sponsors of fans is a beneficial consequence. Again, the teams that do not wear a red shirt should evaluate if the costs of changing is higher or lower than the increase in performance. For teams that are new established the costs of having red or another colored shirt is probably the same. Therefore it would be recommended to choose the color red as the home shirt color.

The use of the different matching methods also gives different effects. Considering every estimated model with the matching method, the nearest neighbor matching method gave the most significant results followed by stratification matching, kernel matching and radius matching respectively. No value judgment can be made on which model to use, preferences should drive the choice for a particular matching model. However, a remark has to be made here. When propensity score matching is used, a data analysts could choose between several matching methods, but all these matching methods have different results in terms of size and significance. Only the direction of the effects is the same in all methods. Still analysts could choose for the model which gives the most favorable results. The problem is that one cannot know what the real and most reliable matching method is. I therefore recommend individuals who use PSM to analyze average treatment effects to check more than one variant of the PSM methods.

## 4.2 Limitations

As in every research, there also exist some limitations to the analysis conducted in this paper. The largest problem in the paper was the absence of a budget variable. For successfully construct the counterfactual no other factors but shirt color should drive the outcome. Although the control variables in this paper try to take over the function of a budget variable, it is not certain if this omitted variable is biasing the results.

Next to this, also the assumption is made that all players have the same additional effects. So an injury of an important player during a season or the dismissal of a coach mid-season at a particular team will happen just as many times in other teams. In practice this assumption will not be realistic, but the bias is minimal because many teams are included in the dataset that played many seasons.

Also a possible unrealistic assumption about the number of residents in home cities is made. Because no data on number of residents when entering the Eredivisie for the first time is available for the Eredivisie teams, the data on the current number of resident in the home cities are used. The assumption is then that the distribution of residents stayed equal during the years.

Finally, next to the wish of more data on controlling variables, the number of teams in the dataset is also limited. If more teams could be compared, a more reliable judgment on the effect of shirt color could be made. To solve this problem, a larger competition than the Eredivisie could be analyzed or the data on teams in the second division of Holland could be included. However, the data on this second division in Holland was not found or unavailable. When having more teams in the dataset, the color categories could also be larger and clustering would not even be necessary. This would cause the results to be more reliable than they are now. If I had more resources and time available, probably this problem could be solved.

## 4.3 Future research

For future research I recommended a more extensive dataset in terms of control variables, but also in terms of the number of teams. Next to this, if one would dig into all the matches played in the Eredivisie, the exact number of matches played by teams wearing a particular color can be counted. In this way the performance when wearing the home shirts can be better compared to the performance when wearing another colored shirt.

# 5. Conclusion

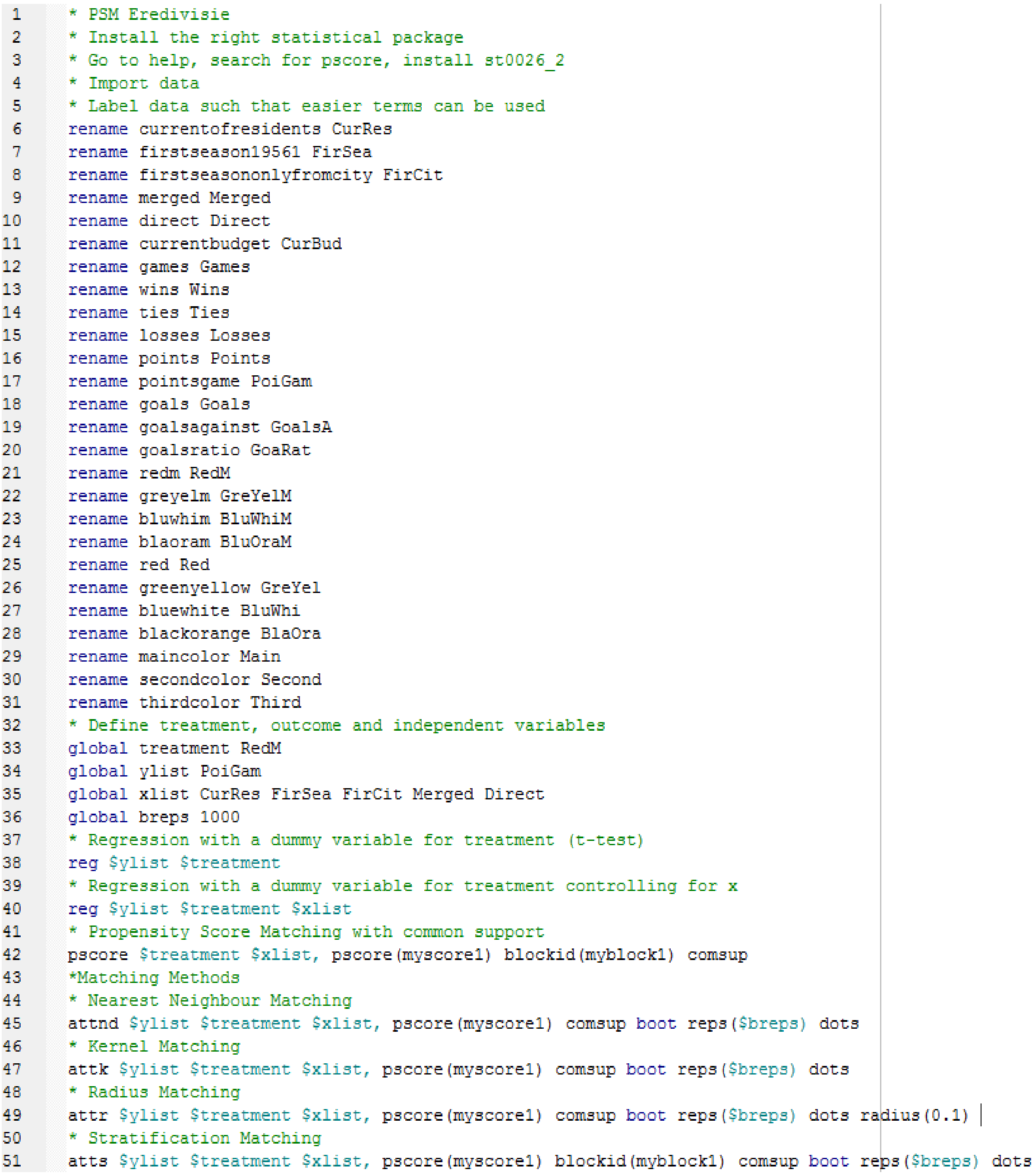
In this paper I have researched the effects of different colored shirts on the performance of football teams in the Eredivisie. The long term discussion on the effects of shirt colors does not give a clear answer to the question where there is an effect. This paper contributes to the ongoing discussion and argues that there exists an advantage for red colored team in getting points or scoring goals. For the Dutch Eredivisie a similar analysis is never been done as far as I know. In that aspect, this research steps into a whole new region that is not often subject to analyzing. Different propensity score matching methods are also used in this paper to find the average treatment effects of shirt colors. It is recommended for teams to change their base colors to red since it yields an advantage over other colored teams. New established teams could enhance their results already upfront by choosing the color red as the major color in their home shirts. Probably the effect of red colored shirts can be assigned to the event that red frightens or seems aggressive to opponents.

# 6. References

- Angrist, J.D. (2008), Treatment effects. In: S.N. Durlauf & L.E. Blume (eds.), *The New Palgrave Dictionairy of Economics*, 2, Palgrave Macmillan.
- Attrill, M.J., Gresty, K.S., Hill, R.A. & Barton, R.A. (2008), Red shirt colour is associated with long-term team success in English football, *Journal of Sports Sciences*, 26, (6), 577-582.
- Bellman, R. (1956), Dynamic programming and Lagrange multipliers. *Proc Natl Acad Sci USA*, 42, (10), 767-769.
- Caliendo, M. & Kopeinig, S. (2008), Some practical guidance for the implementation of propensity score matching, *Journal of Economics Surveys*, 22, (1), 31-72.
- Cramer, J.S. (2003), The origins and development of the logit model. *Logit models from economics and other fields*, 149-158.
- Dijkstrsa, P.D. & Preenen, P.T.Y. (2008), No Effect of Blue on Winning Contests in Judo, *Proceedings: Sciences*, 275, (1639), 1157-1162.
- Guan, W. (2003). From the help desk: bootstrapped standard errors. *The Stata Journal*, *3*(1), 71-80.
- Hill, R.A. & Barton, R.A. (2005), Red enhances human performance in contests, *Nature*, 435, (7040), 293.
- Kaya, N. & Epps, H.H. (2004), Relationship between color and emotion: A study of college students. *College Student Journal*, 38, (3), 396-405.
- Kocher, M.G., Sutter, M. (2008), Shirt color and team performance in football. In: Andersson/Ayton/Schmidt (eds.): *Myths and facts about football: the economics and psychology of the world's greatest sport*, Cambridge Scholars Pub., Cambridge: 125-130.
- Lechner, M. (2001). A note on the common support problem in applied evaluation studies. *Univ. of St. Gallen Economics, Disc. Paper*, *1*.
- Lechner, M. (1999), *Identification and Estimation of Causal Effects of Multiple Treatments Under the Conditional Independence Assumption* (pp. 43-58). Physuca-Verlag HD.
- Lewis, D. (1981), Ordering semantics and premise semantics for counterfactuals. *Journal of Philosophical Logic*, 10, (2), 217-234.
- Piatti, M., Savage, D.A. & Torgler, B. (2012), The red mist? Red shirts, success and team sports. *Sport in Society*, 15, (9), 1209-1227.
- Rosenbaum, P.R. & Rubin, D.B. (1983), The central role of the propensity score in observational studies for causal effects, *Biometrika*, 70, (1), 41-55.
- Veltman, F. (2005), Making Counterfactual Assumptions. *Journal of Semantics*, 22, (2), 159-180.
- Valdez, P. & Mehrabian, A. (1994), Effects of color on emotions. *Journal of Experimental Psychology*, 123, (4), 394-409.

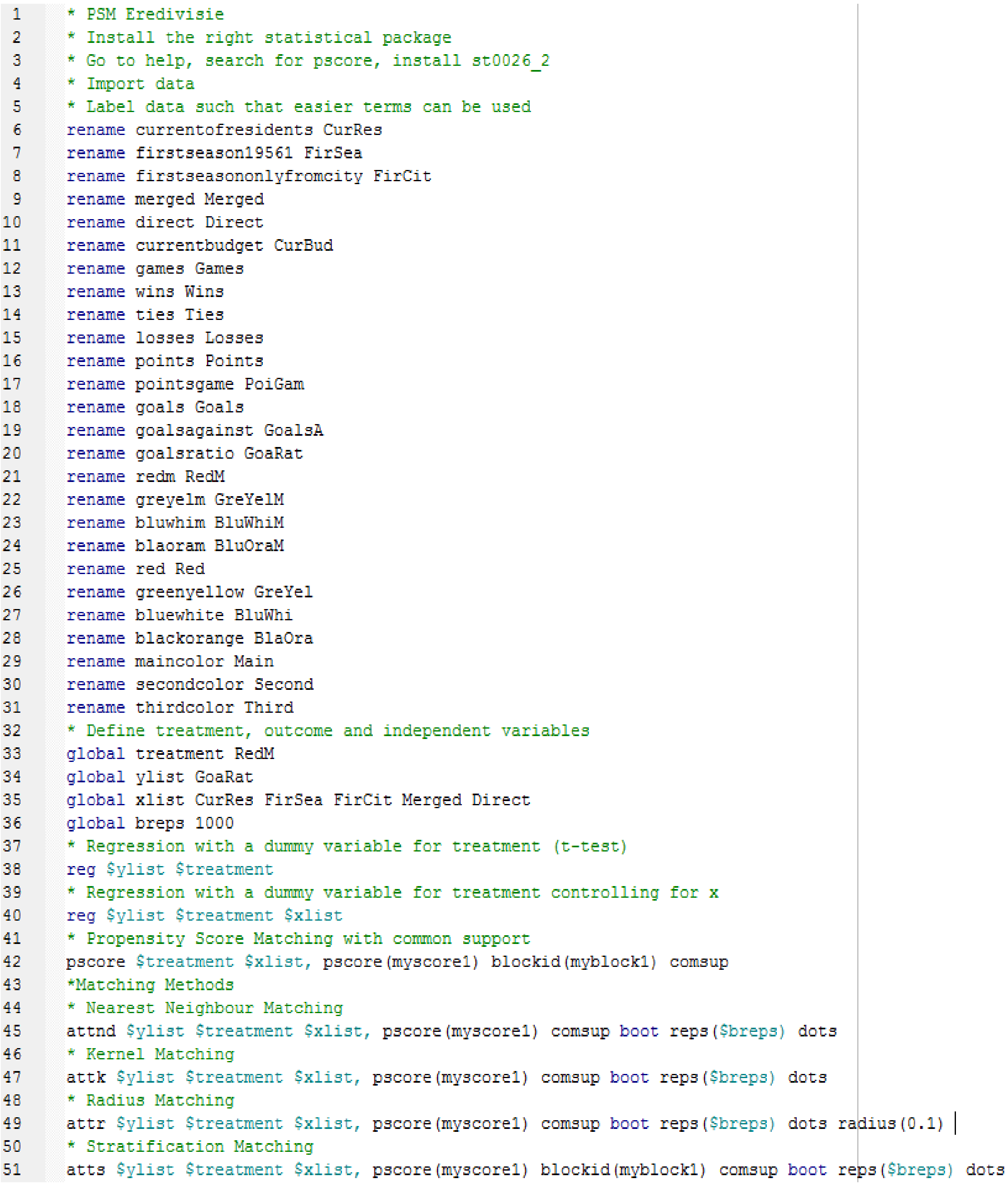
# 7. Appendix

## 7.1 Example of do-file used in the analysis



```
* PSM Eredivisie
* Install the right statistical package
* Go to help, search for pscore, install st0026_2
* Import data
* Label data such that easier terms can be used
rename currentofresidents CurRes
rename firstseason19561 FirSea
rename firstseasononlyfromcity FirCit
rename merged Merged
rename direct Direct
rename currentbudget CurBud
rename games Games
rename wins Wins
rename ties Ties
rename losses Losses
rename points Points
rename pointsgame PoiGam
rename goals Goals
rename goalsagainst GoalsA
rename goalsratio GoaRat
rename redm RedM
rename greyelm GreYelM
rename bluwhim BluWhiM
rename blaoram BluOraM
rename red Red
rename greenyellow GreYel
rename bluewhite BluWhi
rename blackorange BlaOra
rename maincolor Main
rename secondcolor Second
rename thirdcolor Third
* Define treatment, outcome and independent variables
global treatment RedM
global ylist PoiGam
global xlist CurRes FirSea FirCit Merged Direct
global breps 1000
* Regression with a dummy variable for treatment (t-test)
reg $ylist $treatment
* Regression with a dummy variable for treatment controlling for x
reg $ylist $treatment $xlist
* Propensity Score Matching with common support
pscore $treatment $xlist, pscore(myscore1) blockid(myblock1) comsup
*Matching Methods
* Nearest Neighbour Matching
attnd $ylist $treatment $xlist, pscore(myscore1) comsup boot reps($breps) dots
* Kernel Matching
attk $ylist $treatment $xlist, pscore(myscore1) comsup boot reps($breps) dots
* Radius Matching
attr $ylist $treatment $xlist, pscore(myscore1) comsup boot reps($breps) dots radius(0.1)
* Stratification Matching
atts $ylist $treatment $xlist, pscore(myscore1) blockid(myblock1) comsup boot reps($breps) dots
```

## 7.2 Example of do-file used in the analysis



```
* PSM Eredivisie
* Install the right statistical package
* Go to help, search for pscore, install st0026_2
* Import data
* Label data such that easier terms can be used
rename currentofresidents CurRes
rename firstseason19561 FirSea
rename firstseasononlyfromcity FirCit
rename merged Merged
rename direct Direct
rename currentbudget CurBud
rename games Games
rename wins Wins
rename ties Ties
rename losses Losses
rename points Points
rename pointsgame PoiGam
rename goals Goals
rename goalsagainst GoalsA
rename goalsratio GoaRat
rename redm RedM
rename greyelm GreYelM
rename bluwhim BluWhiM
rename blaoram BluOraM
rename red Red
rename greenyellow GreYel
rename bluewhite BluWhi
rename blackorange BlaOra
rename maincolor Main
rename secondcolor Second
rename thirdcolor Third
* Define treatment, outcome and independent variables
global treatment RedM
global ylist GoaRat
global xlist CurRes FirSea FirCit Merged Direct
global breps 1000
* Regression with a dummy variable for treatment (t-test)
reg $ylist $treatment
* Regression with a dummy variable for treatment controlling for x
reg $ylist $treatment $xlist
* Propensity Score Matching with common support
pscore $treatment $xlist, pscore(myscore1) blockid(myblock1) comsup
*Matching Methods
* Nearest Neighbour Matching
attnd $ylist $treatment $xlist, pscore(myscore1) comsup boot reps($breps) dots
* Kernel Matching
attk $ylist $treatment $xlist, pscore(myscore1) comsup boot reps($breps) dots
* Radius Matching
attr $ylist $treatment $xlist, pscore(myscore1) comsup boot reps($breps) dots radius(0.1)
* Stratification Matching
atts $ylist $treatment $xlist, pscore(myscore1) blockid(myblock1) comsup boot reps($breps) dots
```