ERASMUS UNIVERSITY ROTTERDAM

Erasmus School of Economics

Bachelor Thesis Industrial Dynamics and Strategy

Racial Labour market discrimination: The wage difference between football players in Italy

Name student: Jochem Hazelaar

Student ID number: 455659

Supervisor: Francesco Principe

Second assessor: Sam Hoey

Date final version: 29-01-2020

ABSTRACT:

This research investigates racial discrimination in the labour market. Specifically, the difference in wage between white and black football players in the Italian Serie A. The expectation is that white players make more money than their black peers, while their performance characteristics are the same. This is tested with a dataset containing 469 players in a period from 2010 to 2015. In total there are 1585 observations. This data comes from different websites containing information about football players, like their age and different performance characteristics (goals scored and assists). From the ordinary least squared regressions with fixed effects and multiple matching methods follows that it is likely that black players earn less money, but the difference is not significantly different from zero. In conclusion, there is no hard evidence for wage discrimination in the Italian Serie A but there is a possibility that players get discriminated against in other parts of the labour market than the wage. This is in line with research in different football competitions.

I. Introduction

'I have a dream that my four children will one day live in a nation where they will not be judged by the colour of their skin, but by the content of their character' (Nobel Media AB, 2019). However, this dream of Martin Luther King Jr. did not come true yet. The law and public opinion show the awareness of and the willingness to get rid of discrimination, like Martin Luther King Jr. wanted years ago. For instance, in 1969 a law to eliminate all forms of racial discrimination internationally was implemented by the United Nations (United Nations Human Rights, 2019). This law prohibits treating a group or one person worse than others because of who a person is (Human Rights Guide, 2019). This law does not mean that discrimination does not occur anymore. For example, there is still a lack of diversity in the top of corporate companies. These positions are occupied by mainly white men (Brooksworth, 2019). Besides, women get paid less and have higher job insecurity compared to men in the same position (Siddique, 2018). The corporate business is not the only part of the society where discrimination still occurs.

Besides corporate businesses, racial discrimination occurs within the education system. In Australia, 40 per cent of the students experience discrimination from their fellow students (McGowan, 2019). Furthermore, in the health care sector minority groups, experience discrimination regarding their skin colour. In the United States, premature death and illness occur more often among minority groups compared with the dominant group of white people (Tello, 2017). Discrimination is also a problem in the sports world. Hence, female football players get earn less money than their male peers in almost every country despite lawsuits and strikes of women (Lloyd-Hughes, 2019). Unequal pay is not limited to gender, but origin is also a reason to discriminate for chairmen or supporters. A recent study shows that white NBA players have a lower salary than black NBA players while they contribute the same to the team (Wen, 2018). The opposite holds for the football industry where discrimination targets the black players mostly. A few years ago, fans of Zenit St. Petersburg started a petition to sack a black player just because of his skin colour (Telegraph Sport, 2012). Discrimination in sports is not constrained to gender or race, because 40 per cent of the athletes feel intimidated or unsafe because of their sexuality (Prosser, 2019). So, there is a lot of work to do to eliminate discrimination in the world.

The awareness of when and where discrimination should continue to grow before it can be eliminated. The examples above show what is already known but a lot is still unknown as well. This study will try to expand this literature by investigating wage differences based on the race of people. Racial discrimination is treating a group, or a person differently based on skin colour, culture, nationality, or citizenship. (Citizens advice, n.d.). It can have several reasons, for example, a stereotype, negative feelings held by someone, or narrow-mindedness for instance (Ontario Human Rights Commission, n.d.). Further, the definition of wage discrimination is the unequal pay of two people in the same function, with the same performance, but different individual characteristics (gender or sexuality). Wage discrimination occurs based on the same reasons as racial discrimination (Encyclopedia, 2019).

This study utilises sports and football to investigate discrimination because in football it is easy to observe performance and assign the performance characteristics to an individual player. Discrimination is, therefore, easier to investigate within the sports industry than the manufacturing or service sector (Rosen & Sanderson, 2001). There is less discussion about who scored a goal or gave the assist than who made a product within a manufacturing company. Which makes it easy to estimate the performance of a single person. This leads to the following research question:

Do black football players in the Italian Serie A earn less than white football players in the same competition while their performance is the same?

This research will contribute to the literature with discrimination in sports in Europe as a topic. Most studies in this area use samples from the United States and mostly data about the NBA or NFL. Szymanski's (2000) study on discrimination in the English professional football leagues is an exception and the leading with data from Europe. This paper will be an addition to Szymanski's study but will assess discrimination in football leagues in a different manner. Other studies use team-level data to investigate discrimination. This research will use individual-level data. Besides contributing to research with as topic discrimination in sports, this paper will add to research on discrimination in the labour market in general because the sports labour market is a good proxy for other parts of the economy (Kahn, 2000).

The next section discusses the existing research about discrimination in the labour market. Starting with the first theory on discrimination and works towards more recent research which updates this theory. The second part of the literature review consists of a review of research on the topic of (wage) discrimination in sports. Secondly, the data section discusses the data sources and the modifications made to investigate the data properly, like the wage variable. The data section contains a description of the descriptive statistics as well. The methodology section explains the different methods used in this research. The formula and the process of an ordinary least squared regression and the matching process are explained in greater detail. Afterwards, the results are presented. Starting with the results of the regression analysis and secondly the results of the different matching methods and the meaning of these results. Lastly, a conclusion and a discussion contain the main results, conclusions, limitations of the research and suggestions for further research.

II. Theoretical Framework

Becker (1957) established the first economic theory including racial discrimination. In his book, the economics of discrimination, he defines discrimination in a neoclassical way. An individual has a taste of discrimination when he is willing to pay to be associated with a kind of character instead of another character, according to Becker. However, this definition is not uncontroversial, because Marshall, among others, doubts certain parts of it. Becker's definition, for example, assumes discrimination as a physical phenomenon which is not covenanted with reality (Marshall, 1974). Besides, Phelps (1972) states that discrimination follows the belief of employers that one type of workers is more qualified based on non-economic characteristics. Research into the sports labour market provides broader definitions of discrimination. Frick (2007), for example, defines discrimination as a sense that managers or spectators prefer players of a certain origin. This definition drops the part of the grounds of these beliefs. Rosen and Sanderson (2001) also state that discrimination in sports is due to a preference of employers, employees and fans. This research uses the definition of the last two types of research because the designs of these papers are like this one.

Becker concludes that discrimination will not exist in the long run. Firms can discriminate but it will come at the cost of higher salaries for the preferred type of worker. Which type of worker this is, depends on the beliefs of the managers or stakeholders. Firms hiring the group which get discriminated against will, because of the wage differences, outperform the firm who discriminate and have a higher profit. This means, that in the long run the discriminating firms will leave the market and discrimination will be eliminated. The second reason for the vanishing of discrimination, in the long run, is the wage mechanism. The lower wage of the discriminated group will attract firms to hire them and, therefore, push the wage up until the wage gap vanishes. These two phenomena will eliminate wage discrimination in the long run according to Becker's theory (Becker, 1957).

This theory has been influencing and a good starting point, however, the assumptions of the model are not likely to hold in the real economy. Becker's neoclassical theory is the starting point of multiple other pieces of research with as a common goal to estimate the effects of discrimination on the labour market. Aigner and Cain (1977) built a statistical model and relax some of the assumptions made by Becker. There is a wage gap between white workers and black workers, but it is not clear whether discrimination is the cause of it. According to another statistical model, the wage gap between whites and blacks may be due to discrimination. However, only when there are posted wage offers in the vacancies. Workers know their salary before they apply for a job in these situations and, therefore, black workers apply for jobs with a wage low enough for white workers to not apply. Because black individuals know that employers will choose a white worker above a black worker when they can choose between the two types. Whether this statistical model hold is doubtful as well because of the assumptions (Lang, Manove, & Dickens, 2005).

The statistical models are, thus, not clear if discrimination exists in the labour market in the first place and secondly not conclusive if discrimination is persistent in the long run if it exists. The statistical models all confirm the wage gap between white workers and black workers. According to Arrow (1998), it is not possible to explain the phenomenon of discrimination in the labour market with neoclassical economic theory in the first place. Arrow uses rational choice theory to explain discrimination. In short, the rational choice theory assumes that individuals make choices to achieve the optimal benefit of satisfaction for themselves (Ganti, 2019). The theory suggests that individuals discriminate against black workers when they see whites outperforming blacks even when this has nothing to do with race. Maybe do white individuals have a higher level of education or ability. It is not possible to observe somebody's level of education, but it is easy to see whether someone is black or not. This is called statistical discrimination. Besides the rational choice theory, it is also possible to explain discrimination in terms of social interactions and networks. Via social interactions, social segregation occurs in the economy and employers will gain social benefits from hiring one type of workers instead of the other. When these costs of the higher wage bill are lower than the social benefits from discrimination, it will be profitable for employers (Arrow, 1998).

Besides other philosophies to explain discrimination, there are also different explanations for the wage gap between races (Lang, Manove, & Dickens, 2005). Arrow (1998) argued that races can differ in their ability which can cause discrimination. Carneiro, Heckman and Masterov (2005) even state that ability is the foremost reason for the racial wage gap. They use premarket factors, like early schooling results, to separate racial discrimination from a difference in skills. These premarket factors explain a significant part of the wage difference between whites and blacks and show a deficit environment for blacks. It is not possible to use later schooling results because black individuals get confronted with stereotypes when they grow up, which biases them to perform less in tests because they expect to be rewarded in a lesser amount than white students due to the stereotypes. These stereotypes and the following bias are considered as a reason for the wage gap as well (Carneiro, Heckman, & Masterov, 2005). When we go back to the ability as a cause for the difference in wage between races, we see that the difference in ability explains two-third of the racial wage gap using premarket factors as determinants for skills (Neal & Johnson, 1996). A third alternative reason for the wage gap between white workers and black workers is a difference in productivity. Research shows a shortfall in productivity of black employees of 3.3 per cent compared to white employees. This partly explains the wage gap between races, however, the same research finds a preference for white individuals in 56 per cent of the firms, which indicates discrimination (Bowlus & Eckstein, 2002). This preference is visible in a field experiment which looks at call-backs from applications. In this experiment, there is a significant difference between white and black applicants. Black applicants must search for a job twice as long and get fewer call-backs. This experiment shows that there is discrimination in the labour market against blacks but does not say anything about wage (Pager, Bonikowski, & Western, 2009).

These types of research methods are also inconclusive about where the wage gap has his roots. Like the statistical models, other analyses suffer from some drawbacks, which may bias the results. As earlier stated, it is hard to test the appropriateness of the statistical methods and theories, because of the assumptions they make. It is difficult to tell whether the assumptions hold in the real economy (Stiglitz, 1973). Secondly, it is nearly impossible to observe all performance characteristics and take these into account. Even when it is doable observe the whole performance, it is even harder to assign these characteristics to one individual worker in most settings (Pager, Bonikowski, & Western, 2009). Carneiro, Heckman and Masterov (2005) try to proxy the characteristics with early schooling results, which can be effective, however, the proxies are still incomplete and do not fully represent the performance characteristics of individuals. The sports labour market is likely to solve this problem. It is easier to investigate the supply- and demand side in sports, because it is easier to assign performance to an individual athlete but also the preferences of the employers are better visible (Rosen & Sanderson, 2001). However, the sports labour market is just a proxy for the whole labour market, which makes it a good indication for other sectors but it is not extrapolatable towards other labour market sector directly (Kahn, The Sports Business as a Labor Market Laboratory, 2000).

Sports labour market research uses multiple techniques to examine (wage) discrimination. The most common method is an ordinary least squared regression. To investigate racial wage discrimination, researchers use an athlete's salary as the dependent variable and a race dummy variable as an indicator of discrimination. Different performance and individual characteristics are the control variables. Omitted variable bias (OVB) is likely to influence the results of regressions because of unobserved characteristics influencing an athlete's salary. Which makes it difficult to give a useful conclusion (Kahn, 1991). This is the same problem as in the manufacturing or service sector. However, the regressions give a first indication of what the effect of race on salary could be. In a regression analysis, Christiano (1988) finds a discrimination coefficient of 17 per cent against white baseball players. Which means that be a white baseball player earn less money than a black player purely based on their race. In basketball, early research finds an insignificant discrimination coefficient. There is discrimination against black players, but it is not significantly different from zero (Rockwood & Asher, 1976). A third and more recent paper uses a different method, namely a probit regressions, to estimate discrimination in the American football league (NFL). It shows the presence of hiring discrimination against black players, which results in fewer wins for the discriminating teams. The expected compensation for black NFL players is also 1,2 per cent lower than the expected compensation for white NFL players. Despite that the effects are barely significant, it shows wage discrimination (Conlin & Emerson, 2005). Wilson and Ying (2003) use a two-stage least squared regression to see whether there is discrimination against foreign players in the five major football leagues in Europe (English Premier League, German Bundesliga, Spanish La Liga, French Ligue 1 and Italian Serie A). They find little evidence for discrimination based on nationality. Frick (2007) even finds a wage bonus for foreign players compared to German players in the German Bundesliga. However, the careers of Eastern European players are shorter than the careers of players from other nationalities. So, wage discrimination is no issue, but discrimination could exist in the hiring or firing part of the labour market.

Research into (wage) discrimination in sports is inconclusive, just like labour market research in other sectors. Two possible explanations are the different sports because all sports have different characteristics and the variety of research methods. In contrast to Wilson and Ying (2003), one of the most influencing papers finds discrimination in the football labour market. Szymanski (2000) performs a market test on the English professional football leagues. The results are like Becker's prediction of discrimination in the short run. The results show that the employers (teams) with a taste of discrimination will make less profit than non-discriminating employers because of the wage premium they pay to players of the race they prefer. Besides, the total wage bill should reflect the skills and abilities of the players in the teams and, therefore, the place in the league table. When the team has skilled players, they will win more matches and end up higher in the league table. In this way, it is possible to see whether teams pay a wage premium to attract players of the race they prefer. Teams with an above-average share of black players finish higher in the league than their wage bill suggests. These results signal wage discrimination against black players because they have a lower salary but at least the same skills and abilities (Szymanski, 2000).

Most papers use team-level data to assess wage discrimination, however, individual-level data could be more useful to investigate the problem of wage discrimination. This paper will try to do this based on a dataset from the Italian Serie A. The next section discusses the dataset in further detail.

III. Data

This paper investigates discrimination in the labour market of football players in the Italian Serie A using individual-level data. To clarify, the Serie A is the highest professional football league in Italy. The focus is on the wage difference between black football players and white football player while their performance characteristics are the same. The dataset contains 469 outfield players who played in the Serie A in the 2013/2014 season (base season). Goalkeepers are different from outfield players in their characteristics and their performance measurements. So, it is not possible to compare goalkeepers to outfield players and goalkeepers are, therefore, not included in the dataset (Lucifora & Simmons, 2003). The dataset contains information on the players in five following seasons, from the season 2010/2011 until the 2014/2015 season. However, not for every player played in the Serie A those five seasons. This has two main explanations: transfers and the promotion/regulation system. Players transfer to other teams in other countries than Italy or joined a team in the Serie A later than the 2010/2011 season. For those players the data is only available for the years the player played for a Serie A team. When players go from the one Serie A team to another team within the Serie A, only the team variable changes. The second option, the promotion/regulation system, makes that three teams relegate every season to the Serie B (the second league of Italy). Three other teams promote from the Serie B to the Serie A. Unless, players of the relegated teams transfer to a club playing in the

Serie A the next season, they are no longer part of the dataset. In the end, the dataset contains 1585 observations.

The information about these 469 players come from different sources. First, The *La Gazetta Dello Sport*, an Italian newspaper which publishes the annual reports of the clubs represented in the Serie A every year. These reports contain information on the wages of the individual players. The personal characteristics of each player, for example, the position on the pitch, the birth year, and the country of origin are subtracted, mostly, from *transfermarkt.com*. This German website keeps track of all kinds of statistics and news about football. Among these statistics, there is information about individual players and their characteristics and the performance statistics. The information not available on *transfermarkt.com* comes from *soccerways.com*. *Soccerways.com* is a website like *transfermarkt.com* and contains information about players and teams from different competitions. The information about the performance characteristics, like the goals a player scores, the assists a player gives, the minutes a player plays and the yellow- and red cards a player receives are taken from these websites as well.

The most important variable is a dummy variable indicating the skin colour of a player. This dummy variable is one when a player is black and zero when a player is white. However, a player's skin colour is not easy to obtain because it is subjective and different for everyone what is black or not. In the end, discrimination will occur when the manager, the teammates or the fans deem a player to be black, which is subjective as well (Deschamps & De Sousa, 2014). So, to create this dummy variable, a similar method is used as Deschamps and De Sousa (2014) use. A (white) person with some football knowledge judged all 469 outfield players of the base season 2013/2014 by pictures from *transfermarkt.com* or *google.com* and indicated them with a one when he deemed the player to be black and zero when he deemed the player to be white without knowing the wages of the players or their exact performance. Before this person judged all the players by their skin colour, I did the same. The results were different for some players which are shown in table 3 in the appendix and explained in greater detail in the next section.

Table 1 in the appendix shows the statistics of the most important variables and their description. The average wage of a player in the Serie A is circa 875 thousand euros. However, the standard deviation shows a great variance between players (910 thousand euros). The kernel density graph (figure 1) shows a large tail on the right of the distribution of the wages, which suggests a distribution skewed to the right. A few players earn a lot of money (max 6.5 million euros per year) compared to the mean wage. The Kurtosis coefficient shows the same distribution because it is higher than three. To investigate the effect of race on the wage of players, there is a preference for a normal distribution of the wage. To correct for the outliers and make sure the distribution is normal, the natural logarithm of the wage is taken. The Kurtosis coefficient is closer to three for the natural logarithm of the wage and the graph shows a normal distribution (table 2 and figure 2).

The statistics of the race dummy variable show that circa 15 per cent of the players in the Serie A is black in the base season. This static is solely based on the indication the person asked to indicate the skin colour because this judgement is not biased. In the 2013-2014 season, 69 players were black, and 400 players were white. The comparison between the two different black indications shows 18 differences, this was equally divided between the two indications. 50 per cent of these differences were players from Brazil and two-third of the players come from South America. For Europeans or at least Dutch people, it is harder to judge the skin colour of a player from this part of the world than other continents (table 3).

The descriptive statistics of indicate that the average player is 26.4 years old whereby the age varies between 16 and 40 years old. The relationship between wage (natural logarithm) and age is non-linear if you look at the scatterplot (figure 3). To create a better fit, age is squared (age^2) and used to analyse the data. Regarding the positions of the players, almost 40 per cent of the players is a defender and another 40 per cent is a midfielder, and the remaining 20 per cent is an attacker. As mentioned before, the goalkeepers are excluded from the analysis. In the 1352 minutes a player plays on average per season, he scores 1.9 times and gives the key pass leading to a goal (assist) 1.3 times. The goals and assists vary from 0 to 29 goals and from 0 to 14 assists because attackers have more opportunities to score or give an assist. The next part will discuss how these variables are used to analyse the data in greater detail.

IV. Methodology

To estimate the wage discrimination between black and white football players, types of analysis are used. The first one is an ordinary least squared regression (OLS). This method is like other researches and gives a good first impression of wage discrimination. against black players. The coefficient of the dummy variable for the race will be negative when there is discrimination against black football players and positive when white players are subject to discrimination. However, an OLS regression has one major shortcoming. It is only possible to use observable characteristics as control variables, which creates omitted variable bias (OVB). This problem is partly solved by adding fixed effects to the model because fixed effects correct for (un)observed characteristics that stay the same over time. The second part of the analysis uses propensity score matching to minimize the problem of OVB. In this analysis, black and white players are matched based on their performance and the wages of these players are compared. If teams discriminate against black players, the wage will be lower for black players compared to their white matches.

The OLS regression analysis consists of different models and estimations following a general equation:

$$LN(Wage_{it}) = \beta_0 + \beta_1 * D_{it} + \beta_2 * X_{1,it} + \beta_3 * X_{2,it} + \cdots + \beta_k * X_{k-1,it} + \alpha_i + \varepsilon_{it}$$

In this equation $LN(Wage_{it})$ is the natural logarithm of the salary of player i in season t. The constant term (β_0) represents a player's base salary without performing or playing games. The dummy variable for a player's race is the D_{it} and the β_1 is the coefficient indicating discrimination against black or white players. The β_1 coefficient is negative (positive) when white (black) players get paid a higher salary. The X_{it} variables are performance indicators which act as control variables to minimize the OVB. Besides the control variables, some models use fixed effects for the team a player plays and for the season to account for (un)observed time-invariant characteristics (α_i). It is not possible to control for individual fixed effects because the skin colour of the players does not change over time which causes multicollinearity. The error term stands for the unexplained part of the wage of football players and is represented by ε_{it}.

The baseline model contains the discrimination dummy variable, the natural logarithm of wage, all control variables from the dataset and the team, individual fixed effects, and season fixed effects. Some of the control variables probably have a positive influence on the wage, like, the minutes a player played per season, whether the player was captain of the team, the number of goals scored, the number of assists given, the number of fouls committed on the player, and whether the team played in the Champions League (prestigious European club tournament) or Europa League (second level European club tournament). Besides, characteristics that are likely to have a negative effect are part of the baseline model too. Examples are the average fouls a player commits, the amount of yellow- or red cards a player receives and the days a player was injured. Lastly, the baseline model contains a dummy for the position of the players, the age of the player, and of the age^2 as control variables. The control variables are validated in different researches. These variables have a positive or negative effect on the wage of a player (Frick, 2007). However, adding too many control variables can result in a spurious regression. To prevent this from happening, different other models are estimated to check whether these variables are important or not to estimate the discrimination coefficient. It is not possible to remove the OVB completely because there are unobserved characteristics which influence the wage as well. The OLS regression will over- or underestimate the effect of a player's race on his wage and, therefore, probably violate the conditional independence assumption.

Propensity score matching reduces OVB because the functional form is less strict. Within this analysis, the players with a black skin are matched to the white players based on their (performance) characteristics. All individual characteristics represent a certain propensity score and players with the same propensity score are compared to each other. In this way, the wage gap between players with the same performance is shown by the discrimination coefficient. This process follows a few steps. Firstly, the estimation of different logit models. These models follow a general equation:

$$D_{it} = \beta_1 * X_{it,1} + \beta_2 * X_{it,2} + \cdots + \beta_k * X_{it,k} + \varepsilon_{it}$$

Within this equation, the dependent variable (D_{it}) is the dummy for a player's race and the independent variables are the performance indicators and other characteristics from the OLS regression of the previous part. The variables captain and the midfield position are not included because there is no black captain and when the midfield position is included there is the problem of multicollinearity. The βs determine the propensity score together with the dependent variables. The error term is the part not explained by the performance indicators and other characteristics (ε_{it}). The propensity scores of the baseline model from the OLS regressions are used to perform the different matching tests.

The second step is checking the area of common support based on the propensity scores. The scores of both groups are plotted in figure 4. This graph shows a lot of common support only at the right tail, there are no white players with a propensity score over the 0.589, so the black players in the analysis with this score or higher are dropped, because there is no match for these players. For the left tail, the white players with a propensity score lower than 0.016 are dropped because there are no black players to match these players with.

After the test for common support, the black and white players are matched. The three different methods of matching have been applied (nearest neighbour matching, kernel density matching, and radius matching). In nearest neighbour matching, every black player is matched to the white player with the same propensity score. This method has the lowest bias but drops a large part of the sample because there are fewer observations of black players (216) than white players (1,370) in the dataset. This reduces the precision of the analysis. The opposite of nearest neighbour matching is kernel density matching. This method has the highest precision, but also the highest bias. Within this method, every white player is matched to a black player with a weight based on how close the propensity score is. In the last matching method, radius matching, a group of white players within a certain range is compared to a black player. In this analysis, the group of white players within this range differs from two players to five players. The wages between this group and the black player are compared. This method has more bias than nearest neighbour matching but less than kernel density matching and more precision than nearest neighbour matching but less than kernel density matching. The results of the different methods are compared, and this will give an overview of the wage difference between white- and black players. Lastly, the doubly robust estimator is investigated. This method is a combination of matching and an ordinary least squared regression. The doubly robust estimator first computes propensity scores and uses these scores as weights in a weighted regression. This method holds under the conditional independence assumption when the linear functional form of the regression holds or when the propensity scores are estimated correctly.

The effect of being black on the salary of a player is estimated. There will be a negative (positive) coefficient when there is discrimination against black (white) players. This method is better than an OLS regression but there is still a possibility of OVB because matching cannot control for unobservable variables either. The next section will discuss the results of the analysis explained in the methodology.

V. Results

Table 4 and table 5 display the results of the different OLS regressions. The baseline model contains thirteen control variables and the team- and season fixed effects. These variables influence the wage of players differently. The model includes the age^2 to correct for the non-linear relationship between the age of players and their wage. The effect of age is hyperbolic, and the wage increases until the player is 31.6 years old and decreases afterwards. According to the baseline model, the effect of the minutes a player plays per season is basically zero. The enormous number of minutes in a season is a possible explanation for this effect. The third control variable is a dummy variable for whether the player captain is of the team he plays for or not. There are no black captains in the seasons investigated but the effect on the wage of a player is positive which can explain the wage gap between black and white players partly. This is as expected because the captain is an important player and has special abilities. The effects of the number of goals scored by a player, the key passes leading to a goal a player gives (assists) and whether the team a player plays for performs in the Champions League or Europa League are also as expected. These variables all have a positive and significant impact on the wage of players. The fouls made on the player also has a positive effect on the wage of a player, however, this effect is not significant.

According to the baseline model, all control variables positively affect the wage of players. This is unexpected because when a player receives multiple yellow- and red cards, he will be suspended for more often which harms the team's performance. Therefore, it was assumed that the number of yellow- and red cards and the number of fouls committed decreased salaries. An injured player probably misses multiple matches as well and it was, therefore, also assumed that this would decrease the salary of a player. The regression of the baseline model estimates a positive, but insignificant, effect from all variables with an expected negative effect. Which means that they do not differ from zero significantly. This can have multiple reasons, for example, the faith a club has in a player or the performance when the player can play.

The position of a player matters for the wage of players as well. To avoid multicollinearity, the midfielders are excluded from the model. Table 4 shows that the defenders earn almost 9 per cent less than midfielders and attackers earn 20 per cent more money than the midfield players do which is as expected because they have more opportunities to score goals and give assists. Attackers, therefore, have a more important role in the team. The most important variable, the dummy variable for the race of a player is negative (-0.012) but insignificant. Black players earn 1.2% less money than their white peers. However, the effect is insignificant which implies that Italian football teams do not discriminate black players by their wage.

To check whether the fixed effects explain an additional part of the variance, different models with and without the team- and season fixed effects are estimated. Table 4 displays the results of the different regressions. The R-squared and the adjusted R-squared decrease almost by 30 per cent when the team fixed effects are dropped from the baseline model. Not including the team fixed effects increases the discrimination coefficient to 0.060, which means that teams discriminate against white players. Only including the season fixed effects also increase the standard error which lowers the precision of the estimation. Besides, the importance of the team fixed effects, the results also show that it is important to add the season fixed effects to the model. The (adjusted) R-squared decreases by a minimum amount when only the team fixed effects are included. Besides, the discrimination coefficient decreases even further to a salary gap of 3 per cent and the standard error stays the same. Compared to adding no fixed effects, it is better to use fixed effects because it adds a small percentage of explained variance and it changes the discrimination coefficient a fair bit as well. This means that it is better to use the fixed effects regression instead of an ordinary least squared regression without fixed effects. It was not possible to add the player fixed effects to the model because it would cause multicollinearity with the variable of interest. However, it could be that other unobserved individual characteristics that stay the same over time play a role as well. For example, the agent of a player (Ehrhardt & Rodgers, 1988).

The fixed effects are important in examining the wage difference between white football players and black football players. Next, some control variables are removed from the baseline model to examine the importance of these control variables. The minutes a player plays per season has a coefficient of almost zero in the baseline model. The reason for this is probably a large number of minutes in a season. The players possibly play 38 games of 90 minutes, which makes a total of 3420 minutes per season in the league games only. Table 1 shows that the player with the most minutes played 3643 minutes in a season. When the minutes played per season is deleted from the model, the wage gap between black and white players increases to 2.7 per cent and increases the precision. However, the R-squared and the adjusted R-squared decrease which means that the minutes played explains a part of the variance. The dummy for being the captain of the team is dropped as well. This does not change much in the discrimination coefficient, but the explained variance of the data decreases slightly. The fouls a player commits had an unexpected effect on the wage in the baseline model. This is visible in the coefficient of the race dummy variable. The coefficient decreases to -0.037 and the standard error decreases as well. Likewise, the (adjusted) R-squared decreases which means that the fouls a player commits explains a part of the variance. Deleting the yellow- and red cards, lowers the discrimination coefficient and the standard error stays the same. It holds for every variable that the (adjusted) R-squared decreases when the variable is dropped from the model. Dropping the assists given in a season is the last control variable tested. It decreases the R-squared but does not change the discrimination coefficient. Deleting some of these control variables would be in favour of the analysis. However, the individual relationship between those variables and the wage of the player is significant. This means that it is better to use the baseline model than the other models.

To conclude, the baseline model seems to be the best fit for the data with the least bias and the highest explained variance. This would mean that black players have a lower wage compared to their white peers because the coefficient of the race dummy variable is negative. For example, an average player of 26.4 years old who plays 1352 minutes per season in which he scores 1.9 times and gives 1.3 assists as a midfielder, all other variables being zero, he earns 2.218 thousand euros when he is white but earns only 2.216 thousand euros when he is black. This is a difference is made only by the race of a player, however, the analysis can still be subject to OVB, as stated earlier. The next analysis (propensity score matching) tries to minimize this bias.

As described earlier, there is an area of common support. So, it is possible to perform the different types of matching tests using the baseline model of the regression analysis. It is not possible to include fixed effects, for that reason these effects are dropped from the model. Besides, there is no team with a black captain in the dataset which would lead to multicollinearity. The captain variable is, therefore, deleted from the baseline model. The first type of matching is nearest neighbour matching. Table 6 shows the average treatment effect of the treated (ATT). An ATT of -0.030, indicates discrimination against black players. Black players earn three per cent less than their white peers. However, the precision of this analysis is lower than of the other methods. Besides, the effect is not significantly different from zero. The coefficient of the other matching techniques shows a positive coefficient and, therefore, discrimination against white players. For example, the ATT of kernel density is 0.1 and has the highest precision, however, the bias of the coefficient is high as well. Lastly, radius matching based on a different amount of 'neighbours' is performed. This is the middle way between radius matching and nearest neighbour matching. The results in table 6 show an ATT between 0.084 and 0.101 based on the different numbers of white players compared to a black player. The standard error of the analysis with five neighbours is the lowest, which would imply the highest precision. However, no ATT is significant at a 10 per cent significance level and the effects, therefore, do not differ significantly differ from zero as well. In conclusion, it is not possible to state that there is discrimination against black players based on the matching analysis. This may be due to OVB, but it is also possible that teams in the Serie A do not make a difference between black and white players when they determine the wage of a player.

The last analysis of this research is the doubly robust estimator. This is the combination of a regression and propensity score matching. The results in table 6 show an ATT of 0.036, which means that there is discrimination against white players instead of black players. However, this effect is not significant, and the standard error is high compared to the effect itself (0.076). This means that there is no wage discrimination based on the race of players in the Italian Serie A according to the doubly robust estimator.

When all analyses are considered, the conclusion would be that wage discrimination against black players is no issue in the Italian Serie A. The results correspond to the results of other research in football competitions in Europe, like Wilson and Ying's (2003) research. Szymanski (2000) suggests that teams use wage discrimination because teams with an above-average number of black players perform better than a team with more white players with the same overall wage bill. The conclusion is contradictory, but the football competition is different, and the research method is different as well. It is hard to compare the results of different researches for that reason. The next section gives the conclusion and a discussion with the limitations of this research and suggestions for further research.

VI. Conclusion

This research investigates wage discrimination against black football players in the Serie A using the following research question: Do black football players in the Italian Serie A earn less than white football players in the same competition while their performance is the same? The motivation to write this research comes from the news about racism and racial discrimination in sports but also in other sectors of society. Discrimination is a relatively recent topic in research, but discrimination exists for centuries. The first economic theory states that it is impossible to have discrimination in the long run because of competition. This theory is updated later because there seemed to be persistent discrimination in the labour market. Other researchers conclude that the wage difference in the labour market does not follow from discrimination but differences in productivity or schooling. The economic theory is unclear about racial wage discrimination because the performance is hard to observe and assign to an individual employee. Sports is a good proxy for the labour market in general because the performance is easier to assign to an individual and it is easier to observe most of the times. So, discrimination is investigated in different sports and different countries. This research adds to that kind of literature. However, the results of the sports labour market are not easy to extrapolate towards other sectors because there are a lot of differences, but it is a first indication.

The data about the performance of the players come from websites, who keep track of statistics of football players and the information about the wage of players are retained from the year reports of the clubs. To get an indication of the race of the different players, a random person with some football knowledge judged all the players according to a picture on their skin colour. This is in line with other researches using a dummy for skin colour. In the end, around 15 per cent of the players in the league was black in the base season. This data is analysed using a linear regression model and propensity score matching. The baseline regression model uses multiple control variables to get the most reliable outcome. The matching techniques use the same control variables as the regression to estimate the propensity scores. The different matching methods use these propensity scores to estimate the discrimination coefficient. Lastly, the combinations of matching and a linear regression is estimated, the doubly robust estimator.

The results of the baseline model in the OLS estimation show a wage difference of 1.2 per cent between black players and white players in the Serie A. However, this is not a significant effect and can be subject to omitted variable bias. The matching analysis is performed to reduce the OVB even further. From the different methods follows that the discrimination coefficient lies between -0.030 and +0.101. However, the coefficients are not significant, and the standard error is high. Lastly, the doubly robust estimator is positive (0.036) but not significantly different from zero. This would mean that white players earn less money than a black player instead of the other way around. The matching results are subject to OVB as well because in the matching process it is not possible to correct for unobserved characteristics. This is the first drawback of this research.

In the end, racial wage discrimination is not a big problem in the Italian Serie A. There is a small wage gap between black and white players, but it is not significant. So, it could be that black players earn less money than their white peers, but this is not necessarily due to discrimination. The performance of the player has a bigger influence on the wage than the race of the player. This does not mean there is no discrimination in the football labour market in Italy. There is still a possibility that discrimination occurs against black players in different stages of the labour market, for example, in the hiring phase or selling phase. Future research could address these other parts of the labour market.

VII. Discussion

The results do not show a causal relationship between the race dummy variable and a player’s wage because of some serious drawbacks. The first one is the research methods, as described above. Within an OLS regression model and in the matching process it is not possible to correct for unobserved characteristics that bias the relation between the wage and the race of football players. This problem is partly solved by the fixed effects, but this only corrects for the unobservable characteristics that stay the same over time and not for characteristics between seasons. So, these variables are still biasing the analysis which makes it difficult to state that clubs discriminate against black players. To solve this problem, future research can use an instrumental variable or do a random experiment.

The second problem with the analysis is the assessment of the skin colour of a player. The differences between the random person and my judgement are merely based on players from South America and Brazil. So, it is probably better to let a group of people with different cultural backgrounds judge all the players individually and compare the result to address a player as black than just one person. It would be better to do this with Italian people too because most of the managers and directors of the clubs will be Italian in the Serie A. So, future research could focus on this to get a more reliable view on the skin colour of the players.

The last problem is the relatively small number of black players and, therefore, observations in the matching analysis. Which makes it harder to match black and white players to each other. In nearest neighbour matching, the number of matches is restricted to a small amount, which decreases the precision a lot. This problem has a smaller influence on the other types of matching, but a bigger treatment group would help in these situations as well. In future research, it is hard to change this because it is not possible to create more black players to make the treatment group larger. It can be solved by a different research method.

Besides the solutions to the drawbacks from this analysis and the recommendation given in the conclusion section, future research can focus on different competitions and different countries. For example, an individual-level analysis of the English professional football leagues could support the team-level analysis of Szymanski (2000) and analysis in the German Bundesliga could support the research of Frick (2007). More recent data can be used in future research as well to see the changes in discrimination over the years. Racism and discrimination are a lot in the news lately and this could change the way the teams are managed, and the way fans cheer for their clubs.

References

Aigner, D. J., & Cain, G. G. (1977). Statistical Theories of Discrimination in Labor Markets. *ILR Review*, 30(2) 175-187.

Arrow, K. J. (1998). What Has Economics to Say About Racial Discrimination? *The Journal of Economic Perspectives*, 12(2) 91-100.

Becker, G. S. (1957). *The Economics of Discrimination.* Chicago: University of Chicago Press.

Bowlus, A. J., & Eckstein, Z. (2002). Discrimination and Skill Differences in an Equilibrium Search Model. *International Economic Review*, 43(4) 1309-1345.

Brooksworth, S. (2019, November 1). *My time on The Apprentice taught me a lot about black men in business*. Retrieved from The Guardian: https://www.theguardian.com/commentisfree/2019/nov/01/the-apprentice-black-men-business-senior-positions-discrimination

Carneiro, P., Heckman, J. J., & Masterov, D. V. (2005). Labor Market Discrimination and Racial Differences in Premarket Factors. *Journal of Law and Economics*, 48(1) 1-40.

Christiano, K. J. (1988). Salaries and Race in Professional Baseball: Discrimination 10 Years Later. *Sociology of Sports Journal*, 5(2) 136-149.

Citizens advice. (n.d.). *Discrimination because of race*. Retrieved from Citizens advice: https://www.citizensadvice.org.uk/law-and-courts/discrimination/discrimination-because-of-race-religion-or-belief/discrimination-because-of-race/

Conlin, M., & Emerson, P. M. (2005). Discrimination in Hiring versus Retention and Promotion: An Empirical Analysis of Within-Firm Treatment of Players in the NFL. *The Journal of Law, Economics & Organization*, 115-136.

Deschamps, P., & De Sousa, J. (2014). Labor Mobility and Racial Discrimination.

Ehrhardt, C. W., & Rodgers, M. J. (1988). Tightening the defense against offensive sports agents. *Florida State Univesity College of Law*, 16:33.

Encyclopedia. (2019, November 17). *Wage Discrimination*. Retrieved from Encyclopedia.com: https://www.encyclopedia.com/social-sciences-and-law/economics-business-and-labor/labor/wage-discrimination

Equality and Human Rights Commission. (2019, February 27). *Race discrimination*. Retrieved from Equality and Human Rights Commission: https://www.equalityhumanrights.com/en/advice-and-guidance/race-discrimination

Frick, B. (2007). The Football Players' Labor Market: Empirical Evidence From The Major European Leagues. *Scottish Journal of Political Economy*, 422-446.

Ganti, A. (2019, April 1). *Rational Choice Theory*. Retrieved from Investopedia: https://www.investopedia.com/terms/r/rational-choice-theory.asp

Human Rights Guide. (2019, November 18). *Grounds of discrimination*. Retrieved from Discrimination: http://www.cilvektiesibugids.lv/en/themes/discrimination/what-is-discrimination/grounds-of-discrimination

Kahn, L. M. (1991). Discrimination in Professional Sports: a Survey of the Literature. *Industrial & Labor Relations Review*, 44 (3) 395-418.

Kahn, L. M. (2000). The Sports Business as a Labor Market Laboratory. *The Journal of Economic Perspectives*, 14(3) 75-94.

Lang, K., Manove, M., & Dickens, W. T. (2005). Racial Discrimination in Labor Markets with Posted Wage Offers. *The American Economic Review*, 95(4) 1327-1340.

Lloyd-Hughes, F. (2019, June 16). *Vrouwenvoetbal is populairder dan ooit, maar gelijke lonen zijn nog ver te zoeken*. Retrieved from Vice: https://www.vice.com/nl/article/bj95m3/vrouwenvoetbal-is-populairder-dan-ooit-maar-gelijke-lonen-zijn-nog-ver-te-zoeken

Lucifora, C., & Simmons, R. (2003). Superstar Effects in Sports: Evidence from Italian Soccer. *Journal of Sports Economics*, 4(1) 33-35.

Marshall, R. (1974). The economics of racial discrimination: a survey. *Journal of Economic Literature 12.3*, 849-871.

McGowan, M. (2019, August 26). *Racism study finds one in the three school students are victims of discrimination*. Retrieved from The Guardian: https://www.theguardian.com/australia-news/2019/aug/27/racism-study-finds-one-in-three-school-students-are-victims-of-discrimination

Neal, D. A., & Johnson, W. R. (1996). The Role of Premarket Factors in Black-White Wage Differences. *Journal of Political Economy*, 104(5) 869-895.

Nobel Media AB. (2019, November 18). *I Have a Dream...* Retrieved from NobelPrize.org: https://www.nobelprize.org/prizes/peace/1964/king/26141-i-have-a-dream/

Ontario Human Rights Commission. (n.d.). *Examples of racial discrimination (fact sheet)*. Retrieved from OHRC: http://www.ohrc.on.ca/en/examples-racial-discrimination-fact-sheet

Pager, D., Bonikowski, B., & Western, B. (2009). Discrimination in a Low-Wage Labor Market: A Field Experiment. *American Sociological Review*, 74 777-799.

Phelps, E. S. (1972). The Statistical Theory of Racism and Sexism. *The American Economic Review*, 62(4) 659-661.

Prosser, C. (2019, May 5). *LGBT discrimination in sport highlighted after a week in the headlines*. Retrieved from ABC news: https://www.abc.net.au/news/2019-05-05/lgbt-discrimination-in-sport-highlighted-in-inclusion-study/11073598

Rockwood, C. E., & Asher, E. (1976). Racial Discrimination in Professional Basketball Revisited. *American Economist*, 20(1) 59-64.

Rosen, S., & Sanderson, A. (2001). Labor Markets in Professional Sports. *The Economic Journal*, 111(469) F47-F68.

Siddique, H. (2018, September 13). *Workplace gender discrimination remains rife, survey finds*. Retrieved from The Guardian: https://www.theguardian.com/world/2018/sep/13/workplace-gender-discrimination-remains-rife-survey-finds

Stiglitz, J. E. (1973). Approaches to the Economics of Discrimination. *The American Economic Revie*, 63(2) 287-295.

Szymanski, S. (2000). A Market Test for Discrimination in the English Professional Soccer Leagues. *Journal of Political Economy*, 108(3) 590-603.

Telegraph Sport. (2012, December 17). *Zenit St Petersburg fans petition club to sack black players and demand all-white heterosexual team*. Retrieved from The Telegraph: https://www.telegraph.co.uk/sport/football/competitions/champions-league/9750764/Zenit-St-Petersburg-fans-petition-club-to-sack-black-players-and-demand-all-white-heterosexual-team.html

Tello, M. (2017, January 16). *Racism and discrimination in health care: Providers and patients*. Retrieved from Harvard Health Publishing: https://www.health.harvard.edu/blog/racism-discrimination-health-care-providers-patients-2017011611015

United Nations Human Rights. (2019, November 18). *International Convention on the Elimination of All Forms of Racial Discrimination*. Retrieved from United Nations Human Rights - Office of the High Commissioner: https://www.ohchr.org/en/professionalinterest/pages/cerd.aspx

Wen, R. (2018). Does Racial Discrimination Exist Within the NBA? An Analysis Based on Salary-per-contribution. *Social Science Quarterly*, 933-944.

Wilson, D. P., & Ying, Y.-H. (2003). Nationality Preferences for Labour in the International Football Industry. *Applied Economics*, 35(14) 1551-1559.

Tables and Figures

Table 1: descriptive statistics and explanation variables

Variable		Variable explaination	Mean	St. Deviation	Min.	Max.	Observations
Id_name		Identification of a player on number					
Season		Indication for the season (13 is season 2013-2014)					
Wage		Yearly wage in thousand euro	875.126	911.679	30	6500	1586
LN(Wage)		Natural logarithm of the wage	6.384	0.872	3.401	8.780	1586
Black Indication		Dummy indicating skin colour	0.136	0.343	0	1	1586
Black Indication (myself)		Dummy indicating skin colour	0.133	0.340	0	1	1586
Age		Age of a player	26.606	4.230	16	40	1586
Role	Defender	Dummy for position	0.402	0.490	0	1	1586
	Midfielder	Dummy for position	0.399	0.490	0	1	1586
	Forward	Dummy for position	0.199	0.400	0	1	1586
Minutes Played		Minutes played per season per player	1352.152	1067.862	0	3643	1586
Goal		Number of goals per season per player	1.935	3.503	0	29	1586
Assist		Number of assists per season per player	1.282	2.041	0	14	1586

Table 2: descriptive Statistics skewness wage

Variable	Variance	Skewness	Kurtosis
Wage (2013-2014)	769477.8	2.811	12.683
LN(Wage) (2013-2014)	0.807	0.158	3.201
Wage (all seasons)	831158.1	2.495	10.302
LN(Wage) (all seasons)	0.760	0.111	3.242

Table 3: differences in black indication

	All players	Brazilian players	Italian players	Uruguayan players	Chile players	Moroccan players	Belgian players	Portuguese players
I said black	9	5 (55.56%)	2 (22.22%)	0 (0%)	0 (0%)	0 (0%)	1 (11.11%)	1 (11.11%)
Same	451	27	205	14	5	0	2	3
The other person said black	9	4 (44.44%)	1 (11.11%)	2 (22.22%)	1 (11.11%)	1 (11.11%)	0 (0%)	0 (0%)
Total differences	18	9 (50%)	3 (33.33%)	2 (22.22%)	1 (11.11%)	1 (11.11%)	1 (11.11%)	1 (11.11%)
Total	469	36	208	16	6	1	3	4

Table 4: regressions with different fixed effects

	Baseline	Year fixed effects	Team fixed effects	No fixed effects
LN(Wage)				
Black	-0.012	0.060	-0.030	0.045
	(0.054)	(0.073)	(0.054)	(0.073)
Age	0.348***	0.248***	0.358***	0.253***
	(0.057)	(0.080)	(0.058)	(0.081)
Age^2	-0.005***	-0.003**	-0.006***	-0.004
	(0.001)	(0.001)	(0.001)	(0.001)
Minutes Played	0.0000574**	0.000044	0.0000618**	0.0000494
	(0.000)	(0.000)	(0.000)	(0.000)
Captain	0.196***	0.004	0.215***	0.021
	(0.069)	(0.109)	(0.068)	(0.110)
Goal	0.013**	0.026***	0.013**	0.026***
	(0.006)	(0.008)	(0.006)	(0.008)
Assist	0.034***	0.031***	0.032***	0.030**
	(0.008)	(0.012)	(0.008)	(0.011)
Fouled (average)	0.011	0.032	0.022	0.041
	(0.031)	(0.040)	(0.030)	(0.040)
Fouls (average)	0.005	-0.045	0.013	-0.035
	(0.040)	(0.049)	(0.040)	(0.049)
Yellowcard	0.008	0.003	0.007	0.002
	(0.007)	(0.009)	(0.007)	(0.009)
Redcard	0.006	0.027	0.011	0.030
	(0.031)	(0.048)	(0.031)	(0.048)
CL	0.162*	0.986***	0.176**	0.995***
	(0.086)	(0.058)	(0.085)	(0.058)
EL	0.109*	0.438***	0.106*	0.438***
	(0.060)	(0.058)	(0.060)	(0.058)
Daysout	0.027	0.045*	0.025	0.043*
	(0.018)	(0.025)	(0.018)	(0.024)
Defender	-0.088**	-0.143***	-0.077**	-0.134*
	(0.039)	(0.053)	(0.039)	(0.053)
Midfielder	Omitted	Omitted	Omitted	Omitted
Forward	0.205***	0.122*	0.214***	0.125*
	(0.059)	(0.072)	(0.059)	(0.072)
Constant	0.837	2.186**	0.615	2.040*
	(0.786)	(1.108)	(0.808)	(1.124)
Team Fixed Effects	Yes	No	Yes	No
Year Fixed Effects	Yes	Yes	No	No
R^2	0.735	0.488	0.732	0.485
Adjusted-R^2	0.718	0.475	0.717	0.474
Number of observations	742	742	742	742

*significant at 10% level **significant at 5% level ***significant at 1% level

Table 5: Regressions with fixed effects

	Baseline	No minutes played	No captain	No fouls	No yellow card	No red card	No assists
LN(Wage)							
Black	-0.012	-0.027	-0.026	-0.037	-0.031	-0.023	-0.012
	(0.054)	(0.052)	(0.052)	(0.049)	(0.052)	(0.053)	(0.054)
Age	0.348***	0.350***	0.316***	0.304***	0.317***	0.314***	0.325***
	(0.057)	(0.576)	(0.056)	(0.050)	(0.056)	(0.057)	(0.058)
Age^2	-0.005***	-0.006***	-0.005***	-0.005***	-0.005***	-0.005***	-0.005***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Minutes Played	0.0000574 **						
	(0.000)						
Captain	0.196***	0.200***					
	(0.069)	(0.069)					
Goal	0.013**	0.018***	0.020***	0.022***	0.022***	0.022***	0.032***
	(0.006)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)
Assist	0.034***	0.039***	0.039***	0.038***	0.040***	0.039***	
	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)	
Fouled (average)	0.011	0.018	0.023	0.019	0.025	0.021	0.047
	(0.031)	(0.031)	(0.031)	(0.029)	(0.031)	(0.031)	(0.030)
Fouls (average)	0.005	0.001	-0.003		0.037	0.039	0.014
	(0.040)	(0.040)	(0.040)		(0.035)	(0.035)	(0.034)
Yellowcard	0.008	0.014	0.015**	0.015***			
	(0.007)	(0.006)	(0.006)	(0.005)			
Redcard	0.006	0.011	0.013	0.018	0.020		
	(0.031)	(0.031)	(0.032)	(0.029)	(0.031)		
CL	0.162*	0.175**	0.179**	0.133	0.182**	0.202**	0.217**
	(0.086)	(0.086)	(0.086)	(0.081)	(0.086)	(0.087)	(0.090)
EL	0.109*	0.116*	0.120**	0.115**	0.126**	0.137**	0.138**
	(0.060)	(0.060)	(0.060)	(0.056)	(0.059)	(0.059)	(0.060)
Daysout	0.027	0.025	0.028	0.020	0.025	0.025	0.021
	(0.018)	(0.018)	(0.018)	(0.018)	(0.010)	(0.018)	(0.018)
Defender	-0.088**	-0.075*	-0.079**	-0.072*	-0.070*	-0.077*	-0.115***
	(0.039)	(0.039)	(0.039)	(0.037)	(0.039)	(0.040)	(0.039)
Midfielder	Omitted	Omitted	Omitted	Omitted	Omitted	Omitted	Omitted
Forward	0.205***	0.179***	0.172***	0.169***	0.136**	0132**	0.090
	(0.059)	(0.056)	(0.056)	(0.052)	(0.054)	(0.053)	(0.055)
Constant	0.837	0.646	1.063	1.212*	1.040	1.086	0.947
	(0.786)	(0.798)	(0.782)	(0.702)	(0.784)	(0.789)	(0.809)
R^2	0.735	0.733	0.731	0.724	0.729	0.726	0.717
Adjusted R^2	0.718	0.717	0.715	0.709	0.713	0.710	0.702
Number of observations	742	742	742	829	742	744	744

*significant at 10% level **significant at 5% level ***significant at 1% level

Table 6: matching results.

		Nearest neighbour matching	Kernel density matching	Radius matching (2 neighbours)	Radius matching (3 neighbours)	Radius matching (4 neighbours)	Radius matching (5 neighbours)	Doubly robust estimator
Baseline model	ATE	0.051	0.147*	0.099	0.131*	0.189*	0.166	0.284***
		(0.123)	(0.080)	(0.081)	(0.075)	(0.100)	(0.113)	(0.077)
	ATT	-0.030	0.100	0.094	0.101	0.095	0.084	0.036
		(0.094)	(0.098)	(0.090)	(0.161)	(0.065)	(0.052)	(0.076)
Observations		742	742	742	742	742	742	742

*significant at 10% level **significant at 5% level ***significant at 1% level

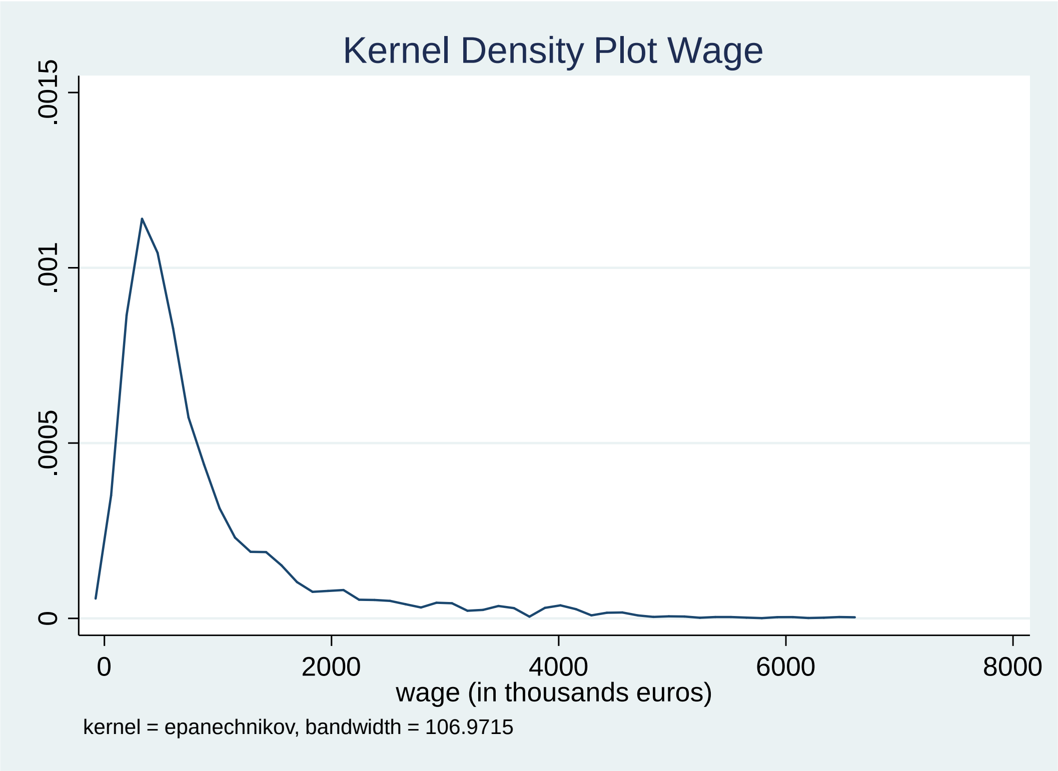

Figure 1: kernel density plot of the wage

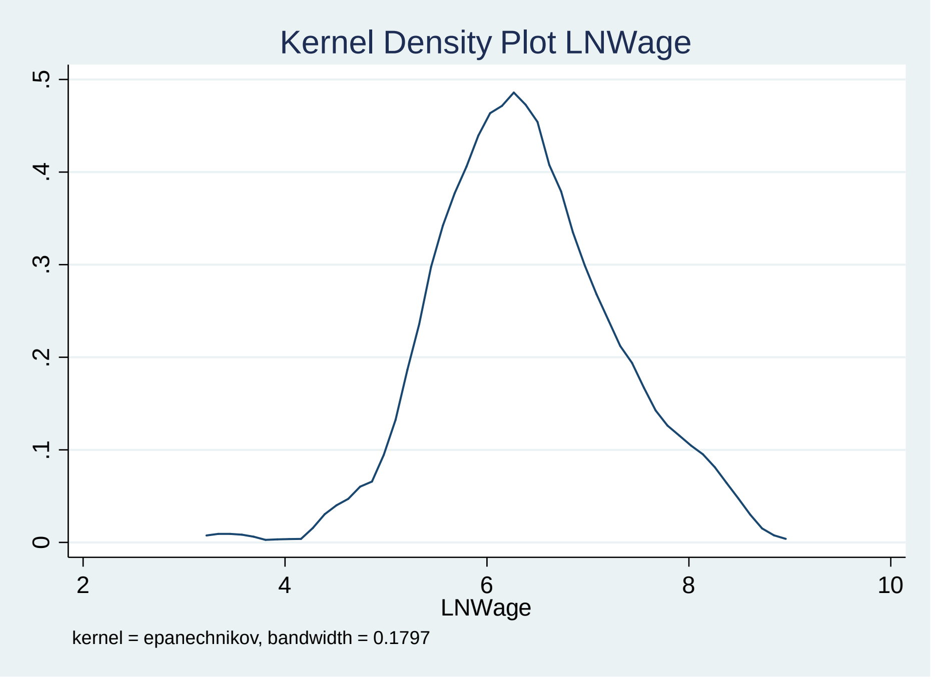

Figure 2: kernel density plot of the natural logarithm of the wage

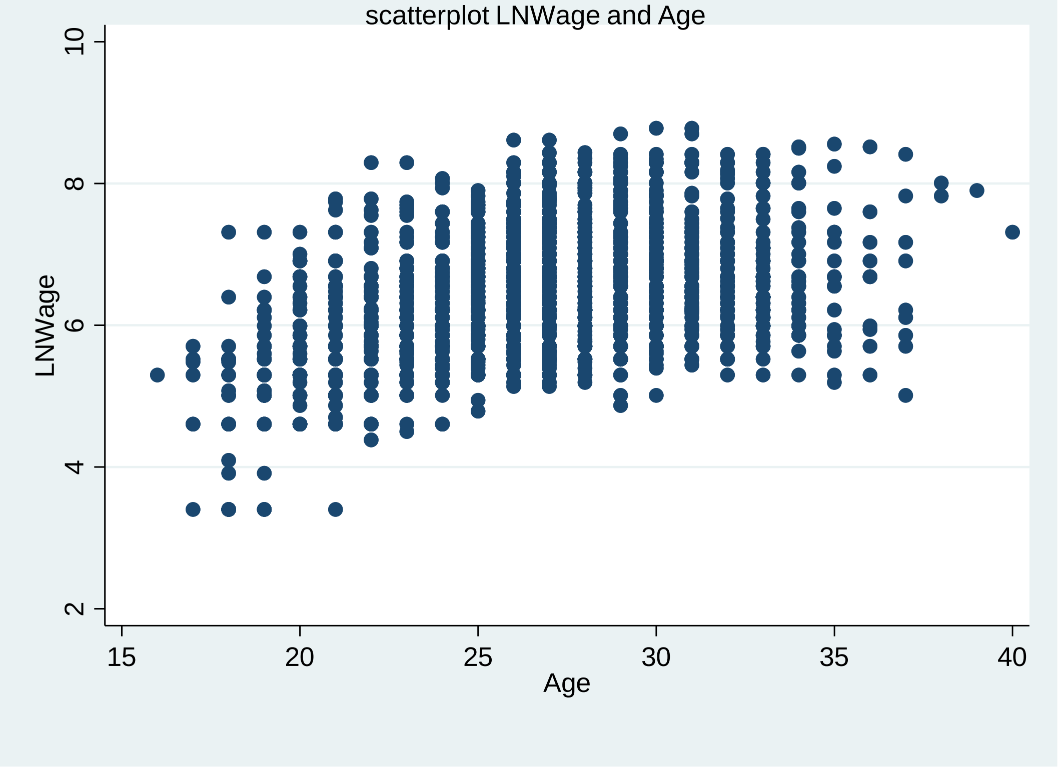

Figure A.3: scatterplot lnWage and Age.

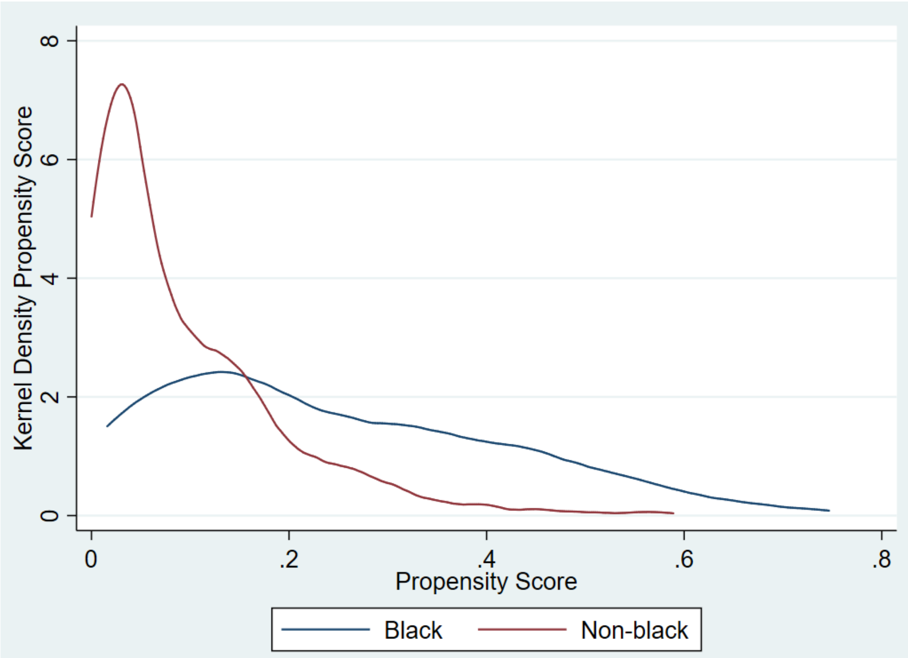

Figure 4: common support graph