Explaining English Premier League Wage Differentials by Nationality

Hanna Ginzburg

Abstract

Do English players command higher wages than non-English players in the Premier League? This paper aims to identify and analyze the existence of nationality-based employer discrimination in football wage-setting, using multiple linear regression models built on data from SoFifa.com. The first regression incorporates eight distinct covariates, along with interaction terms and team and position fixed effects, that may causally influence wage. The second includes the same covariates but assigns overall rating as the dependent variable. Subsequent analysis informs the conclusion that employer and statistical discrimination against non-English nationalities is unlikely; the wage differential is more likely a result of a combination of factors, particularly lower ability standards, and team-specific and player-specific differences. The homegrown player quota actually lowers employment standards for English players, resulting in a negative relationship between English and overall player rating (used as a proxy for ability). English players see a gradient in wage, too; they earn more for a one-unit increase in overall rating and for playing a single position, versus multiple positions, when compared to non-English players.

Introduction

Since the inception of professional football associations in the United Kingdom, there has been contentious debate regarding the degree to which they operate as non-discriminatory, just, and even tangibly meritorious bodies. This paper aims to explore the inner-workings of the English Premier League, colloquially known as the Premiership within the United Kingdom and the country's primary football association. In 1992, the top 22 clubs, or teams, within the English Football League (EFL) split off to form this more elite league; in 1995, membership was cut to 20 clubs and remains so for the 2017-18 season Campbell (2002). Seasons run from August to May, with teams playing 38 matches each (they play each team twice, once at home and once away). It functions as a corporation in which the 20 member football clubs, or teams, act as shareholders and are allocated revenue *2006 FA Premier League* (n.d.). The Premier League has the highest revenue of any football league in the world, earning about £4.5bn in 2016-17. In the same season, each team received a flat participation fee of £35,301,989 and additional payments for TV broadcasts, commercial rights, and a theoretical measure of "merit" based on final league standings (League, 2017). It functions on a system of promotion and relegation with the English Football League Championship, the second-highest league in the U.K. The three lowest ranked teams in the Premier League are relegated to the Championship, and the top two teams from the Championship are promoted to the Premier League, with an additional team promoted after a series of play-offs involving the third, fourth, fifth and sixth placed clubs. EPL teams can compete for four top-tier domestic football competitions: the Football League Cup, the Premier League, the FA (Football Association) Cup and the FA Community Shield *Goal.com* (n.d.).

Fortunately, there exists an abundance of statistics on Premier League football players' abilities and characteristics, as well as an archive of models created to better understand revenue, wage, or performance based on these characteristics. This paper aims to explore a specific relationship within the Premier League, which,

to the best of my knowledge, has not yet been directly explored – that of nationality and wage. I specifically look to the gradient in wage as it affects English players, for whom the Premier League is most likely a native association.

The underlying motivation for this analysis results from initial-stage, exploratory bar charts that I created to plot wage in millions of USD on the y-axis (as well as in log form), against nationality groups on the x-axis. These show a substantial gap between English and non-English groups (see figure 1). Indeed, the figure suggests that non-English players appear to command higher overall and average wages of $29,657 per week than English players. My duty is to figure out, first, if this difference is significant, and second, if it is driven by discrimination in the form of preferential treatment for international players, or if it is a result of other factors (like the homegrown player quota). I use a cutoff level of 5% to determine significance, Stata to conduct my analyses, and Tableau to create visualizations.

Beginning with the first question, I draw on data from SoFifa.com to build an initial multiple regression model and conduct an appropriate analysis. With English included among all other covariates, the first regression model is able to explain 97.8% of the variation in wage; ultimately, English is negatively correlated with wage, though it is not significant at a 5% level. The most statistically significant predictors of wage are overall rating, age, and specific teams and positions – with a gradient for English players in the form of differing returns to higher ratings and certain positions. Importantly, teams do not pay English players a notably different wage than their non-English counterparts.

I then build a second multivariate regression model, which is able to explain 74.3% of the variation in overall rating, to explore the role of the homegrown player rule in wage-setting. The homegrown player rule, created and implemented by the English Football Association, stipulates that each Premier League football club must consist of at least eight English players, defined as those who have played on an English team for at least three years before the age of 21. This qualification is much easier to fulfill for players of English origin, who do not have to permanently relocate to another country at the vulnerable age of 18. In light of this, I hypothesize that players who are English by nationality (as listed on SoFifa.com) might be likely to benefit from this quota in the form of less strict hiring standards on the basis of ability. I use the secondary regression output to determine that there exists a significant negative correlation between English and ability, as measured by overall rating. Conditional on EPL employment, an English player with the same profile as a non-English player can expect to both receive a lower wage *and* be assigned a lower ability rating. The evidence thus suggests that the wage differential between English and non-English players is a result of the homegrown player rule, among other possible factors. This negates the possibility of employer discrimination and instead supports the importance of ability in determining wage in the Premier League.

Literature Review

According to the Premier League, the organization operates to the overwhelming benefit of the British economy: "Premier League football generated £2.4bn in tax revenues in 2013/14, and contributed £3.4bn to the UK's Gross Value Added." The EPL provides a range of support for the development of English football, including solidarity payments for all EFL and National League clubs, and the funding of pensions, welfare, and medical insurance costs for all EFL players. The league also recently launched an initiative called Primary Stars, which provides free teaching resources in order to inspire primary school children in everything from maths and English to teamwork and sport (League, 2017). Evidence of wage discrimination within the organization's ranks would be disastrous for their image, though two academic papers have already cited the existence of implicit discrimination in the Premier League. One finds that white referees award significantly more yellow cards to non-white players of oppositional identity, and another finds that referees are biased towards their home team. Gallo, Grund, and Reade (2012) Dawson, Dobson, Goddard, and Wilson (2007).

Much of the academic literature available on the topic of the English Premier League builds evaluative

models to analyze the performance of teams or players. One article uses an econometric frontier model to evaluate the football clubs' performances from 1998-1999 to 2002-2003, concluding that efficiency scores of the clubs are mixed but each is dependent in part on the price of players (in other words, their wages). Wage is an important factor in determining the efficiency of clubs, as higher wages necessarily drive up operational costs. The paper finds that "in the elite cluster (Manchester United, Arsenal, and Chelsea), Manchester United scores adequately, reflecting the matching of sporting performance and financial activities, whereas Arsenal is in the middle of the rank, and Chelsea displays the least efficient score" (Barros & Leach, 2006). Teams that pay higher wages must be able to recover higher revenues, in part by performing well on the field and selling large numbers of tickets. Refer to figure 7 to visualize the premium in wage payments to players employed by this "elite cluster." Teams that spend more on wages also tend to hire the most skilled players; "all of the clubs have large wage bills and large squads but will only attract players commensurate with their aspirations. The best go to the best, and success breeds success" (Barros & Leach, 2006).

Another paper identifies differences in performance by country and emphasizes the role of country-specific factors in widening these performance gaps, including culture, demography, and geography. Per-capita wealth and country-specific weather conditions are some of the most important factors in determining success. The authors conclude that "beyond a certain level, greater wealth can be subject to diminishing returns, i.e. performance increases with per-capita wealth at a decreasing rate." And, "the deviation of average temperature from 14 degrees Celsius... has a significant effect on a country's football performance." Size of a country's population, however, has no significant impact on its performance in international football (Ramasamy, 2002). Though this article discovers several important factors driving success in football performance, it does not shed light on equally pressing particulars: whether or not an individual player's performance is dependent on the country he hails from, and whether or not players are paid according to how they perform. My analysis will attempt to fill in some of these gaps by analyzing the wage differential between players of English and non-English origin, to determine the extent to which this can be explained by ability, as captured by overall rating.

Further literature develops various methods of measuring player performance and outcomes. Guanhao et. al employ a Skellam process to represent real-time betting odds for English Premier League games, demonstrating their volatility (Guanhao, Nicholas, & Jianeng, 2016). Dobson et. al develop a dynamic game-theoretic model of optimizing strategic behavior on the part of football teams (Dobson & Goddard, 2010). Arabzad et. al use data envelopment analysis to select and rank the best football players in the league. This information then allows us to identify which teams are able to recruit these top players and whether or not they are paid an appropriate wage for the true value they bring to team outcomes. Arabzad et. al label the top three players as Wayne Rooney, Didier Drogba, and Carlos Tevez (Arabzad, Ghorbani, & Shahin, 2013). Wayne Rooney is from England and currently plays as a forward for Everton F.C with a wage of $134,000; to put this in perspective, the current wage maximum for all English players in the league is $185,000 per week. Didier Drogba is from the Ivory Coast and earns a salary of about $4.7 million per year, or $91,000 per week. He is now retired from international football but is best known for his career at Chelsea, for whom he has scored more goals than any other foreign player. Carlos Tevez is an Argentine forward who earns a salary of about $38 million per year, or $730,000 per week, and has historically played for West Ham and Manchester United.

Elliott and Weedon argue that involvement of foreign players in the English Premier Academy League is actually advantageous, but they base their conclusions in large part off of interviews, as opposed to econometric analyses. They cite one interviewee, an academy coach, who asserts "The argument is that there's too many foreign players, so the English kids don't get a chance, but if the English kids are good enough, they'll get a chance. By raising the bar and by bringing in different sorts of players, better players, the quality will rise" (Elliott & Weedon, 2010).

Fewer articles explicitly treat wage as the object of interest in evaluating policies by professional sports associations. One paper employs a structural empirical model to show that UEFA Financial Fair Play regulations, established in 2009 to prevent professional football clubs from spending more than they earn,

should substantially reduce competition. This should result in lower average wages, without changing average revenues, for each team. Depending on the exact regime, wage to turnover ratios should decline by 8% to 15%. The authors also find that "home advantage matters in English soccer" (Peeters & Szymanski, 2012). Lenten et. al estimate a hedonic model of player market value by determining factors that inform the valuation of professional athletes in cricket's Indian Premier League (IPL). They find both expected results, such as more experienced players receiving higher market bids, and unexpected ones, including the presence of home player bias (Lenten, Geerling, & Kónya, 2012).

While the use of regression modeling to identify significant determinants of player value (as assigned by performance or wage) is not new, to my knowledge, it has never before been applied to the English Premier League in an attempt to identify employer discrimination. This paper employs regression modeling in an effort to begin filling in this gap – though it is by no means a conclusive study.

Data and Descriptive Statistics

The data used in my analysis originate from SoFifa.com, for 720 players employed by various football clubs in the Premier League. It is a cross-sectional set, up to date as of October 2017. The entries are unique on the basis of player name and include values for the dependent variable, wage (in thousands of USD to make the results easily interpretable for American readers), as well as eight covariates. The covariates are English (a binary variable: 1 for a player of English nationality, as listed online, and 0 otherwise), team (in text), age (in years), overall rating (in numerical form), height (in inches), weight (in pounds), market value (last traded price or projected lifetime value, in millions of USD), and position (in integer form, with six values corresponding to each possible position).

English is the constructed independent variable, based on player nationalities as they are reported online, whose effect on wage we attempt to isolate.

Team, used interchangeably with F.C. (football club), is specified on the basis of the knowledge that some football clubs have more money at their disposal to spend on player wages. So, a player for Manchester United F.C. of exactly equal standing as a player for Burnley F.C. may earn a higher wage simply because Manchester United has more revenue at its disposal. Or, a team may place more value on its players and accordingly allot a greater percentage of its revenue toward players' earnings.

Age is included to account for the possibility that players are no longer considered to be in their "prime" after a certain age and thus are projected to perform at less than their potential in subsequent years; their wages may reflect this presumption. Younger players, for example, may be more attractive to football recruiters, since they may be expected to provide value to the team for longer. Alternatively, some recruiters may prioritize more experienced players and are willing to pay a premium for them.

Overall rating is included because players with higher scored abilities, averaged over many categories, may be expected to earn a wage premium. This variable, as sourced from SoFifa.com, is calculated through a fairly complex procedure. It starts with guesswork, as a network of over 9,000 members reviews each player's abilities, watches him play, and helps assign him various ratings. EA Sports calls these members data reviewers, and they are made up of coaches, professional scouts, and season ticket holders. Once this group submits opinions on each player, their feedback is handled by 300 editors, who arrange it into 300 fields and 35 attribute categories, including ball control, shot power, etc. They eventually aggregate these scores into one average "overall" ability rating. I chose to use overall rating as a measure of ability over potential rating after regressing wage on each variable and finding a noticeably higher R squared value in the regression using overall rating. Intuitively, this makes sense, as players should be paid according to their current value, rather than their projected future value.

Height is specified to account for the possibility that taller players have better abilities, inflating their weekly wage. A taller player, for instance, can block more of the net with outstretched limbs; accordingly, they might

make better goalkeepers. Some football recruiters may therefore offer a wage premium to the taller player (or to the shorter player) when choosing between two players that are exactly alike barring their height. This depends on the particular skills the particular recruiter values the most or needs to account for when staffing a football team. Weight is included for a similar reason; staffers may idiosyncratically offer a wage premium to the lighter player when choosing between two that are exactly alike aside from their weight. They might assume that lighter players will be able to run faster and perform better. Or, vice versa — managers might think heavier players garner more force in kicking the ball and this distinction demands a wage premium.

Position takes six distinct values: goalkeeper, striker, multiple positions, defense, forward, and midfield. This variable is relevant because playing in a certain position may be more difficult or demanding; for example, strikers might be compensated at a premium because there is more pressure on them to score goals. Or, multitalented players might be compensated at a higher wage because they can sub in for different players at short notice.

Market value is specified because it is the price that each player's current club paid (or is willing to pay) to purchase him, which should be correlated with the wage offered to the same player. A steep market value should command high wages.

Interestingly, English players tend to be younger, shorter, and lighter on average by comparison to non-English players, and have lower average wages, market values, and overall ratings. Though the nationality gradient colorizing the variables related to physical build is beyond the scope of this paper, my regression models analyze the gradient affecting wage and overall rating. Market values are calculated as arguably more complicated functions of performance, wage, age, and external politics (players or managers can vouch for specific players, and fans' interest and support in specific players can drive up their value) so I will not attempt to more closely evaluate the market value gap by nationality.

Empirical Methodology

The estimation strategy I use to analyze the effect of being an English player on wage is a multiple linear regression model. In my first regression, I include other variables relevant to determining a player's wage in order to isolate the precise causal effect of being English on earnings. When configuring exactly how to include each variable, I plotted the relationship between wage and each predictor variable separately; I noted a curvilinear relationship between wage and age, and wage and overall rating. I thus include age and overall rating as polynomial variables in the equation, specifically in quartic form. I also constructed six fixed effects, or dummy variables, for the six possible levels of position, and 22 fixed effects for team.

So as to answer the research question in the most straightforward way possible, I recoded player nationality into a binary English variable. After running a basic regression and noting a highly significant coefficient on English (as well as a non-ideal R squared value), I included interaction terms for English crossed with all other covariates to see if this omittance may have been artificially making the association between English and wage stronger. The inclusion of interaction terms for each predictor variable did, in fact, resolve at least some of this omitted variable bias. I also include the output for each fixed effect in the regression results in order to highlight the particular teams and positions that are correlated with differences in wage for English players and to quantify their effects. In my first regression, I log both the wage and value variables to make the results interpretable in percentage terms. The resulting regression model is as follows:

Log Expected Wage = β_0 + β_1(English)+ β_2(Age) + β_3(Age^2) + β_4(Age^3) + β_5(Age^4) + β_6(Overall Rating) + β_7(Overall Rating^2) + β_8(Overall Rating^3) + β_9(Overall Rating^4) + β_{10}(Height) + β_{11}(Weight) + β_{12}(Log Market Value) + β_{13}(English*Age) + β_{14}(English*Overall Rating) + β_{15}(English*Height) + β_{16}(English*Weight) + β_{17}(English*Log Value) + Position fixed effects + Team fixed effects + English*Position fixed effects + English*Team fixed effects

The basic coefficient of interest is β_1, which represents how much more an English player can expect to earn per week (in percentage terms) when compared to a non-English player who has the same age, overall rating, height, weight, market value, and position on the field, and who belongs to the same Premier League team. With a unit increase in height, the percentage change in wage for an English player is estimated by $(\beta_0+\beta_1+\beta_{10}+\beta_{15})$*100%, and the percentage change for a non-English player is quantified by $(\beta_0+\beta_{10})$*100%. $(\beta_1+\beta_{15})$*100% thus gives the expected percentage change in wage for an English player who grows by one foot, compared to a non-English player who also grows by one foot, with all else held constant. The same logic applies to proper interpretations of all other interaction terms.

After noting a negative correlation between English and log wage, I build a similar multiple linear regression model to isolate the effect of being English on overall rating, which operates as a proxy for ability in this context. This second model helps explain the second part of my research question: whether or not the wage differential for English players is due to employer discrimination, or some other reason(s). It determines whether or not the gap in overall rating of 3.53 points between English and non-English players (see figure 2) represents a real inferior playing ability on the part of English players. In this case, I do not log the market value variable, to match the format of the dependent variable. The resulting equation is as follows:

Log Expected Wage = β_0 + β_1(English) + β_2(Age) + β_3(Age^2) + β_4(Age^3) + β_5(Age^4) + β_6(Height) + β_7(Weight) + β_8(Market Value) + β_9(English*Age) + β_{10}(English*Height) + β_{11}(English*Weight) + β_{12}(English*Age) + β_{13}(English*Overall Rating) + β_{14}(English*Height) + β_{15}(English*Weight) + β_{16}(English*Market Value) + Position fixed effects + Team fixed effects + English*Position fixed effects + English*Team fixed effects

Figure 8 is included as part of a sensitivity analysis to check the assumptions needed for making inferences from the regressions. The pattern displayed when residuals are plotted against fitted values for the second model does appear to be fairly randomly scattered about 0. On the other hand, the first model, which takes log wage as the dependent variable, looks like it may be missing a chunk of important values; further analysis is needed.

Weaknesses of this methodology can be grouped into two basic forms: possible manifestations of omitted variable bias or attenuation bias. Apart from these, there are issues related to the possibility of extrapolating from these results to evaluate the overall presence of discrimination in the Premier League. For one, the inclusion of a basic dummy variable for English and non-English nationalities does not allow the model to identify a wider gradient that would inform wage discrimination by nationality on a per-country basis.

One of the most salient issues related to my regression specification is the possibility that I fail to identify certain factors crucial in determining wage. Wage may be affected in part by another independent variable, either related or unrelated to the quality of being a non-English player, not captured in my model. If this is true, I cannot conclude that the pure quality of being a non-English player will raise expected wages, even if β_1 is statistically significant. It may be the case that the quality of being a non-English player affects at least one omitted factor that in turn raises expected wages for non-English players, or that the explicit inclusion of specific variables informally captured in the error term would allow for a different, more accurate value of β_1. Those who pick players and assign wages for a particular position, height, weight, or another factor may grant higher wages to those matching their idiosyncratic preferences. But, the positive wage bias shown by some managers toward taller players could, for example, be canceled out or made less apparent in the model by a negative bias by other managers toward taller players. Other omitted variables are easier to pinpoint; for example, an indicator for players that are left-footed or ambidextrous is left out. The data I used also includes a single aggregate measure for ability (overall rating), rather than specifying different values for measures like speed, scoring accuracy, and dribbling ability. The exclusion of any of these factors leaves room for omitted variable bias.

Second, there may be some measurement error within the data I used; it is cross-sectional and does not include

a ton of observations. Even with 500+ it is difficult to attribute a causal effect. A more powerful analysis may draw on panel data with more observations to see how relationships, like the one between player position and wage, change over time. It is likely that certain positions have been more or less valuable at different points in time, over the course of the Premier League's lifespan. Furthermore, the data did not originate from a randomized controlled experiment. I instead had to create a model based on gathered, crowdsourced data from SoFifa.com, the construction of which did not involve random assignment of nationalities or objective calculations. Analysts, club managers, and others often make subjective decisions about players' abilities and market values. This possibility could show up in the form of attenuation bias in the coefficients of some of the constructed model's covariates.

Results

The value of β_1 in the first regression illustrates that an English player can expect to earn a lower weekly wage by 57.7% of a non-English player's wage (p=0.226). Though this result is not statistically significant, adding the model's standard error to the β_1 estimate allows us to quantify the upper and lower bounds of the true parameter value; the resulting range is all negative, [-105.3%, -10.1%]. There does appear to be some negative effect of being English on wage, though it may not be exactly -57.7%. The β_1 estimate in the second regression implies that an English player with the same profile as a non-English player is associated with a lower overall rating by -7.025 points, albeit with a much higher standard error of 11.48, making the upper bound estimate of β_1 positive at 4.455 (p=0.612). The correlations between English and log wage and English and overall rating are likely both negative; but with low corresponding p-values, we cannot definitively ascertain causality.

Applying these results within the context of EPL rules and regulations, we can use the homegrown rule to explain, at least in part, why English players are paid less. Players that are English by nationality (as listed on SoFifa.com) are more likely to benefit from the quota in the form of less strict hiring standards and have a greater chance of being recruited by a Premier League team. EPL hiring managers likely compromise somewhat on overall rating when hiring English players, in order to fulfill the eight-person quota. In other words, a Premier League-employed English player that is the same as a non-English player on all other relevant metrics except wage and ability can be expected to have a lower ability level, measured by overall rating. If EPL wages are determined largely on the basis of ability, it is only fair that players with lower ratings get paid lower wages. And, because English players have lower ratings, they may be allocated lower wages than those of their counterparts. It is unlikely that the story runs the other way, because EPL wages are initially always determined by predicted performance, and one of the best measures of performance is rating.

The models' results further negate the possibility of statistical discrimination (paying English players lower wages simply due to their association with a group of lower average ability rating) through the identification of a gradient for returns to one-unit increases in ability. English players very likely earn 7.84% more in wages than non-English players (p=0.018) when both groups are affected by a one-unit increase in overall rating, all else held constant. This difference is quantified by the interaction coefficient between English and overall rating in the first regression and signifies that English players with higher than average ratings are rewarded with higher wages. In other words, the marginal financial return to overall rating is steeper for English players compared to non-English players.

Shorter and heavier players reap greater expected financial rewards; a one-inch increase in height is associated with lower expected wages by -0.639% (p=0.206) and a one-pound increase in weight is associated with higher expected wages by 0.42% (p=0.584). A higher market value also appears to have a negative influence on expected earnings, by -0.5% (p=0.602). However, these results are so minuscule and statistically insignificant that height, weight, and value more likely have no bearing on wages in the EPL. The effects of age and overall rating on wage are more difficult to interpret because these relationships are not linear. Increases in player age do appear to have some positive influence on wage, perhaps due to the experience players gain over

the years. Increases in overall rating appear to have some negative influence on wage, but the effect is less negative for players of English origin.

The coefficients on position and team fixed effects account for much of the remaining variation in wage. Individuals who play a forward position earn 8.86% more in weekly wages (p=0.0128), while strikers earn 7.84% more (p=0.0065); goalkeepers earn 34% less (p<0.0001) and defenders earn 10.2% less (p=0.0016), relative to those who play multiple positions. Playing in an offensive position (that allows one to score goals) perhaps unsurprisingly pays more than a defensive position, as football matches are evaluated on the basis of goals scored.

By far the greatest amount of earnings variation in the first regression is captured by the team dummy variables, however. Discrepancies by football club in the Premier League are huge, and most are significant at a 5% cutoff. Individuals signed with Arsenal, Bournemouth, Chelsea, Everton, Leicester City, Liverpool, Manchester City, Manchester United, Newcastle United, Southampton, Stoke City, Swansea City, Tottenham Hotspur, West Bromwich Albion, and West Ham United earn more than similarly profiled individuals signed with Brighton and Hove Albion, at a statistically significant level. The coefficient on Chelsea is the largest out of those that are most significant, with individuals earning 48% more in weekly wages than those signed with Brighton and Hove Albion (p<0.0001), after controlling for all other potential sources of variation in the model. Players signed with Burnley, Fulham, Huddersfield Town, Hull City, and Watford all earn significantly less than similarly profiled individuals signed with Brighton and Hove Albion. Among these, the largest and most significant effect is the one attributed to Hull City; individuals who play for Hull City earn expected weekly wages that are 129.2% less than those of Brighton and Hove players (p<0.0001), after controlling for all the same potential sources of variation in the model.

Much of this may also be attributed to ability; as far as related literature is concerned, it is not disputed that the top three teams, Chelsea, Arsenal, and Manchester United, employ the best players. The second regression results corroborate that players signed with Chelsea, Arsenal, and Manchester United are expected to have higher overall ability ratings by 7.907 points (p<0.0001), 8.014 points (p<0.0001), and 9.776 points (p<0.0001), respectively. Wealthy and prestigious teams surely attract more effective players since they can offer more enticing perks (both financial and other).

The lack of significant coefficients on interaction dummy variables for English crossed with each football team further negates the presence of discrimination by teams on the basis of nationality. This means that wages within a particular team do not vary significantly between English and non-English players. Only two English * F.C. interaction coefficients are significant at a 10% level in the second regression – that of English*Manchester City and English*Arsenal (both negative for English players, with p=0.0653) – and we cannot disprove that this was due to pure randomness, out of 21 such coefficients. I thus fail to find condemning evidence of nationality-based employer or statistical discrimination; the wage differences in the Premier League are better explained by the interplay of ability rating, team, and position.

Conclusions

In this paper, I attempted to determine if discrimination against non-English players exists in the Premier League wage market. My main findings indicate that this is not the case, though further analysis is necessary to come to a conclusive verdict. By these regression results only, an English player should not earn a significantly different wage than a non-English player the same on all other relevant metrics included in the model. The observed absolute wage differential must be explained through other means; most likely, players are paid wages according to their ability (rating), the team that employs them, and how much potential goal-scoring value they can bring to the team (their position). Players of English origin are associated both with lower ratings and lower wages; and, English players see higher returns in wage from one-unit increases in rating. This, as well as the lack of significance for most team interaction fixed effects, suggests that English players are simply getting paid earnings relative to the value they bring to their team.

There are several theories that can explain the negative correlation between English players and lower overall rating; the homegrown player rule is one. Though we cannot attribute the effect entirely to a specified quota of eight English players per EPL team, it is certainly plausible to conclude that ability standards for hiring non-English players are more stringent in the Premier League partly as a result of this rule. Greg Dyke, chairman of the Football Association, actually wants to increase the quota to 12 and tweak the definition of homegrown players to those who have played for an English team at least three years before turning 18. If so, it will be interesting to note the effect on ability ratings and wages for English and non-English players.

These results ultimately help inform the impartiality of wage determination in the Premier League, but they do not provide a holistic summary. There may still be wage discrimination on the basis of nationality not picked up in this analysis. The most pertinent next step would be to gather more observations and build a regression model with the inclusion of every nationality represented in the Premier League: one that does not operate under the relatively crude distinction between English and non-English players. Such a model may draw on panel data, to determine how the relationship between variables like position and wage changes over time. It may also warrant the inclusion of further covariates, like more specific measures of player ability and left footed-ness or ambidexterity.

Perhaps more striking than any other finding is the significance of team in determining player wage. If the Premier League, and the FA more broadly, want a more equitable distribution of earnings, both bodies may have to reconsider their existing policies. There may have to be a renewed push to cap wages further, so as to assuage the massive wage differentials within the league – perhaps encapsulated most strongly by the 129.2% gap between players for Hull City and Brighton and Hove Albion.

Appendix

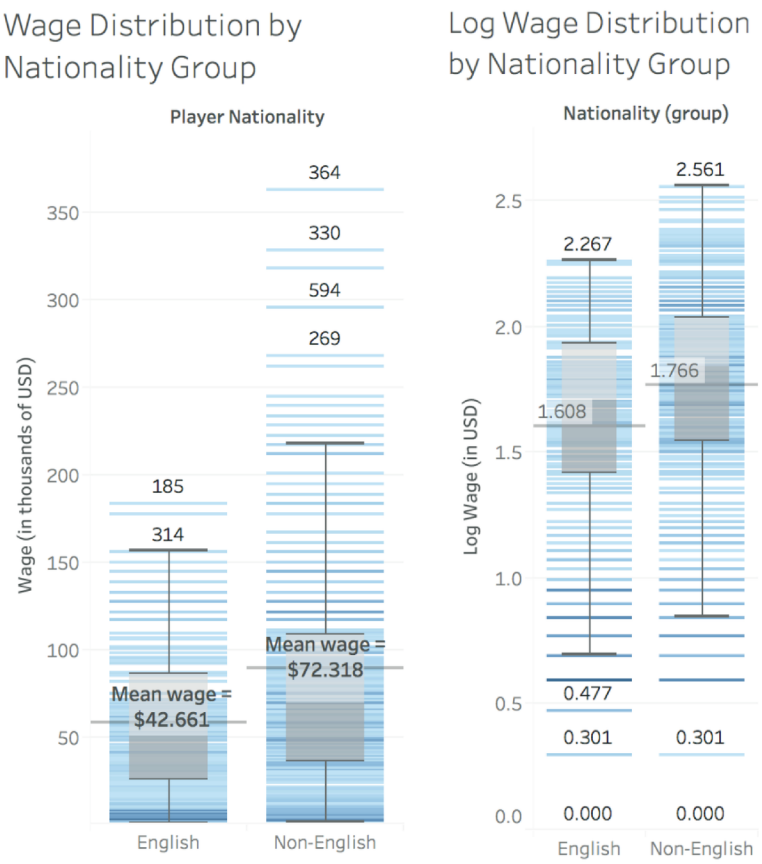

Figure 1: This figure motivates my research question, as it shows a substantial wage gap between English and non-English football players, when comparing both their overall distributions and mean wages. The first suggests that non-English players appear to command higher wages of $29,657 per week than English players. When the dependent variable is log weekly wage, the distributions appear more similar, but a clear disparity still exists.

English	Mean	Std. Dev.	Minu-mum	Maxi-mum	Non-English	Mean	Std. Dev.	Mini-mum	Maxi-mum
Wage (in USD)	45488	41319	2000	185000	Wage (in USD)	77128	61341	4000	364000
Age (in years)	23.39	5.0	16	37	Age (in years)	25.64	4.41	16	38
Overall rating	69.07	8.64	47	86	Overall rating	74.56	7.92	50	90
Height (in inches)	71.38	2.7	65	79	Height (in inches)	72.13	2.84	60	79
Weight (in pounds)	163.76	16.52	115	220	Weight (in pounds)	171.57	17.59	130	223
Market value (in USD)	1450701070416340		9000	728000000	Market value (in USD)	20283420670461 20		7000	672000000

Table 1: Summary statistics for the quantitative predictor and dependent variables included in the models.

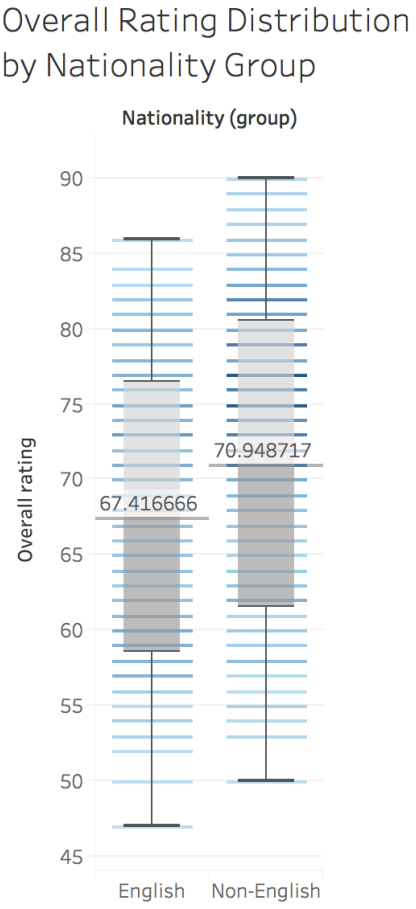

Figure 2: This figure corroborates the story that English players are paid less because they are rated as having lower ability, on average. By plotting the distributions of overall rating by group, we see a lower average rating by 3.53 points for English players.

VARIABLES	(1) Log wage	(2) Overall rating
English	-0.577	-7.025
	(0.476)	(11.48)
Age	2.542***	-51.02***
	(0.694)	(17.99)
Age^2	-0.138***	3.560***
	(0.0407)	(1.060)
Age^3	0.00342***	-0.101***
	(0.00104)	(0.0273)
Age^4	-3.25e-05***	0.00100***
	(9.80e-06)	(0.000259)
Overall rating	-15.01***	
	(1.972)	
Overall $rating^2$	0.322***	
	(0.0419)	
Overall $rating^3$	-0.00301***	
	(0.000392)	
Overall $rating^4$	1.04e-05***	
	(1.36e-06)	
Height	-0.00634	-0.159
	(0.00505)	(0.128)
Weight	0.000835	0.0275
	(0.000829)	(0.0215)
Log market value	-0.00508	
	(0.00974)	
English * Age	-0.00737	-0.0218
	(0.00553)	(0.0860)
English * Overall rating	0.00784**	
	(0.00330)	
English * Height	0.00420	0.113
	(0.00767)	(0.193)
English * Weight	-0.00171	0.00751
	(0.00131)	(0.0317)
English * Value	2.70e-06	-0.00127
	(0.000170)	(0.00556)
English * Forward	0.175**	-1.189
	(0.0711)	(1.799)
English * Striker	0.230***	-4.056**
	(0.0739)	(1.932)
English * Midfield	0.181***	-1.832
	(0.0635)	(1.506)
English * Defense	0.132**	-1.845
	(0.0621)	(1.538)
English * Goalkeeper	0.192**	-3.549*
	(0.0870)	(1.974)
Forward	0.0886**	-1.369
	(0.0356)	(1.053)
Striker	0.0784***	0.991
	(0.0288)	(0.994)
Midfield	-0.0312	-0.526

Figure 3: Main regression results.

	(0.0288)	(0.773)
Defense	-0.102***	-1.386*
	(0.0323)	(0.779)
Goalkeeper	-0.340***	-2.696***
	(0.0468)	(1.033)
Arsenal	0.480***	8.014***
	(0.0535)	(1.163)
Bournemouth	0.259***	1.433
	(0.0630)	(1.716)
Brighton and Hove Albion	-	-
Burnley	0.0325	-0.829
	(0.0530)	(1.154)
Chelsea	0.593***	7.907***
	(0.0513)	(1.311)
Crystal Palace	0.0191	0.846
	(0.0566)	(1.837)
Everton	0.527***	5.135***
	(0.0507)	(1.250)
Fulham	-0.0840	-1.356
	(0.0939)	(1.111)
Huddersfield Town	-0.0971	-0.837
	(0.0761)	(1.418)
Hull City	-1.292***	-1.381
	(0.0588)	(1.220)
Leicester City	0.0973*	4.189***
	(0.0519)	(1.400)
Liverpool	0.565***	6.509***
	(0.0504)	(1.313)
Manchester City	0.624***	9.345***
	(0.0551)	(1.511)
Manchester United	0.654***	9.776***
	(0.0515)	(1.320)
Newcastle United	0.202***	1.717
	(0.0471)	(1.246)
Southampton	0.346***	4.286***
	(0.0489)	(1.387)
Stoke City	0.183***	2.913**
	(0.0503)	(1.299)
Swansea City	0.181***	0.812
	(0.0488)	(1.398)
Tottenham Hotspur	0.374***	7.293***
	(0.0536)	(1.575)
Watford	-0.121	2.618**
	(0.0829)	(1.167)
West Bromwich Albion	0.271***	1.724
	(0.0701)	(1.249)
West Ham United	0.414***	4.426***
	(0.0576)	(1.095)
English * Arsenal	0.0132	-3.651*
	(0.0885)	(1.981)

Figure 4: Main regression results.

English * Bournemouth	-0.0441	0.382
	(0.0775)	(2.186)
English * Brighton and Hove	-	-
English * Burnley	0.0862	1.062
	(0.0751)	(2.000)
English * Chelsea	-0.000545	-2.106
	(0.0952)	(2.104)
English * Crystal Palace	0.0713	0.00489
	(0.0742)	(2.331)
English * Everton	0.0149	0.141
	(0.0742)	(2.089)
English * Fulham	0.186	3.343
	(0.147)	(2.404)
English * Huddersfield Town	0.100	-1.604
	(0.111)	(2.147)
English * Hull City	0.0218	0.628
	(0.103)	(1.968)
English * Leicester City	0.0784	-1.672
	(0.0807)	(2.096)
English * Liverpool	-0.0699	-0.382
	(0.0746)	(2.098)
English * Manchester City	-0.0780	-5.037*
	(0.0928)	(2.733)
English * Manchester United	-0.103	-2.001
	(0.0723)	(2.228)
English * Newcastle United	0.0603	-1.542
	(0.0715)	(2.071)
English * Southampton	-0.0121	-1.971
	(0.0743)	(2.053)
English * Stoke City	-0.00366	-0.203
	(0.0801)	(2.182)
English * Swansea City	0.0660	1.114
	(0.0816)	(2.303)
English * Tottenham Hotspur	-0.00764	-1.372
	(0.0829)	(2.546)
English * Watford	0.0436	0.271
	(0.108)	(1.960)
English * West Bromwich Albion	-0.0939	-0.257
	(0.0879)	(2.049)
English * West Ham United	0.0473	-2.694
	(0.0765)	(1.755)
Market value		0.00118
		(0.00479)
Constant	247.8***	310.5***
	(35.35)	(112.3)
Observations	720	720
R-squared	0.978	0.743

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Figure 5: Main regression results.

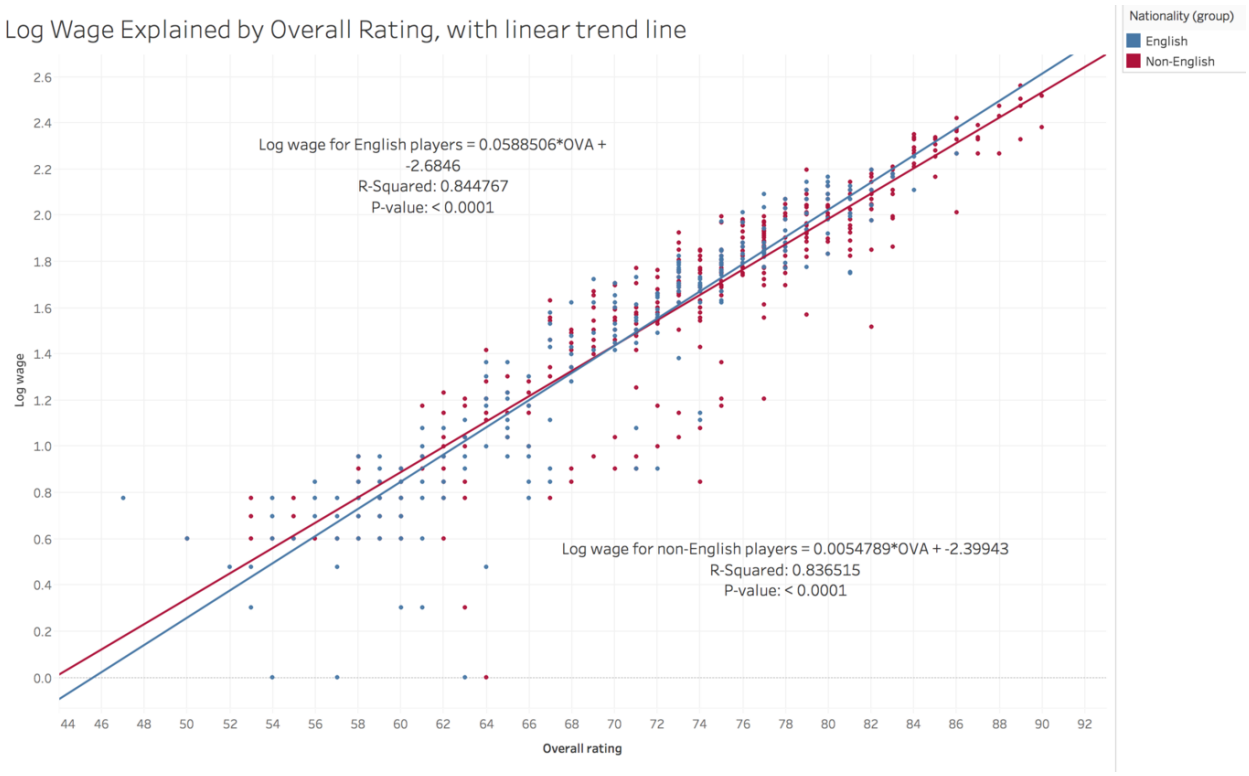

Figure 6: This univariate linear regression shows there is a slight difference in how closely overall rating tracks, or predicts, wage for Premier League football players on the basis of their nationality.

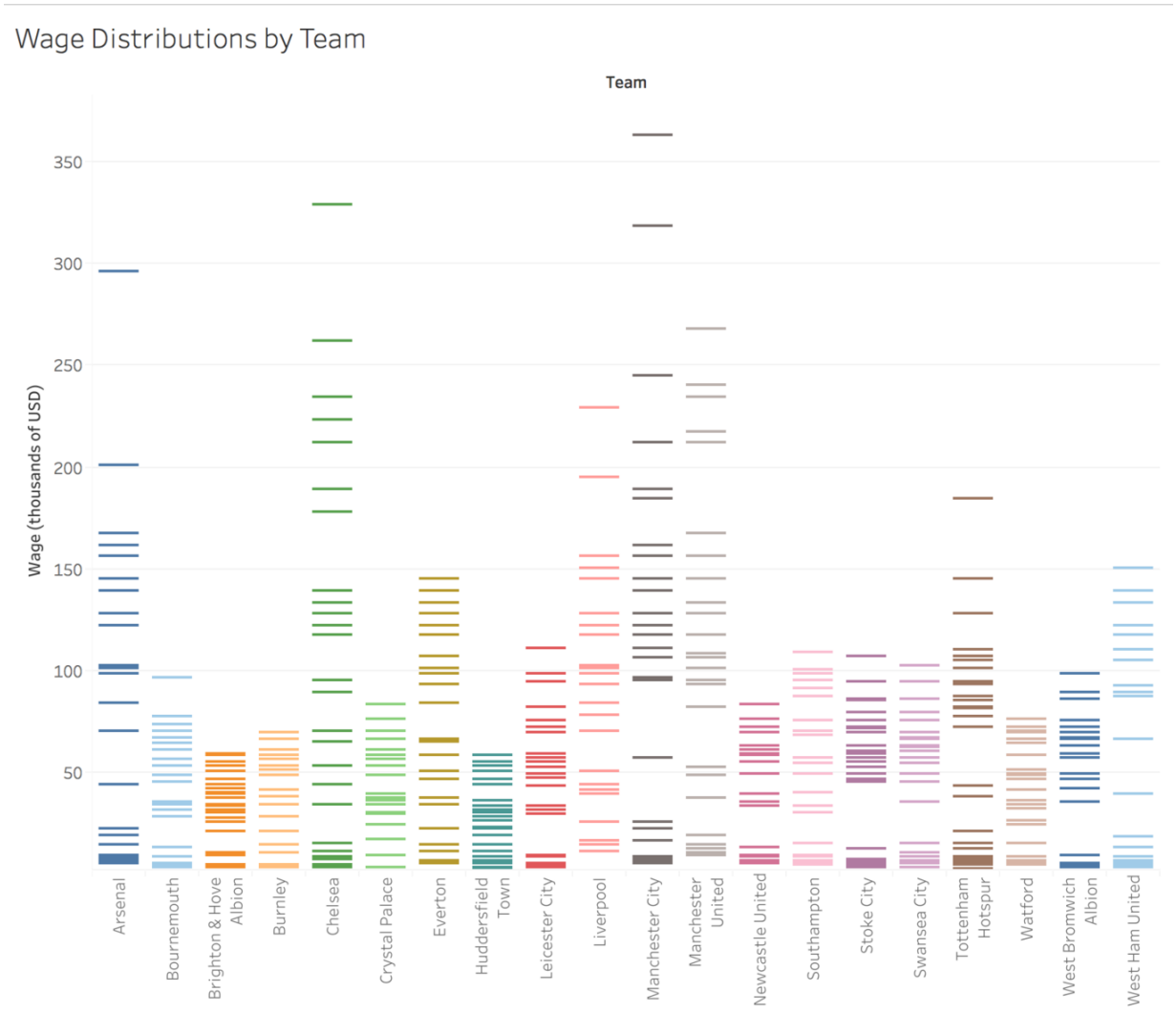

Figure 7: This figure plots wage distributions by team. Clearly, Manchester City F.C. spends the most on its players, with Chelsea and Arsenal close behind - making up the "elite" of the Premier League.

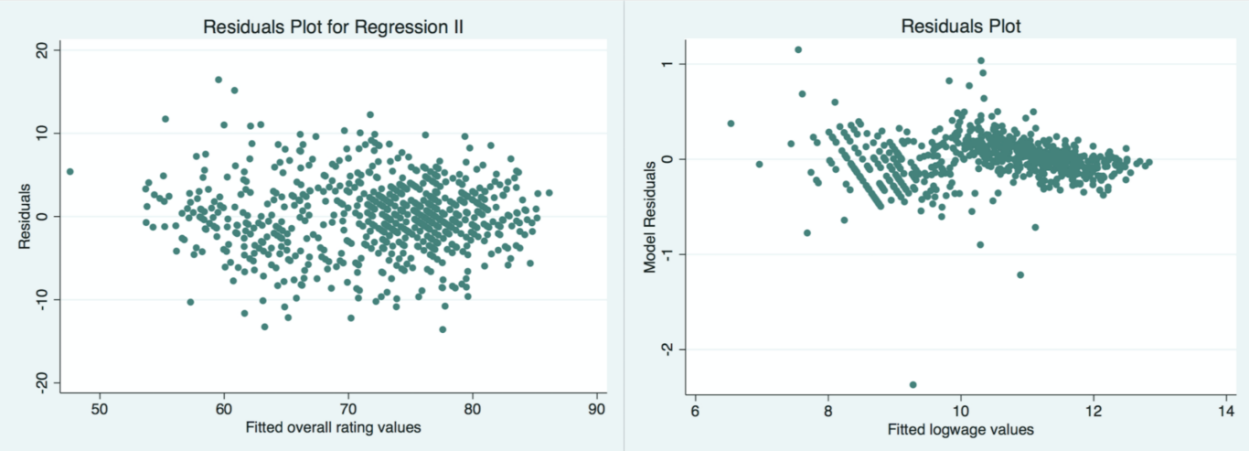

Figure 8: The residuals versus fitted values plots for both regressions show a relatively random pattern around 0. This serves to confirm the validity of each model, though the first model estimating log wage appears to be missing a chunk of important values. Further analysis would be able to confirm whether the variance in residuals is truly random or not.

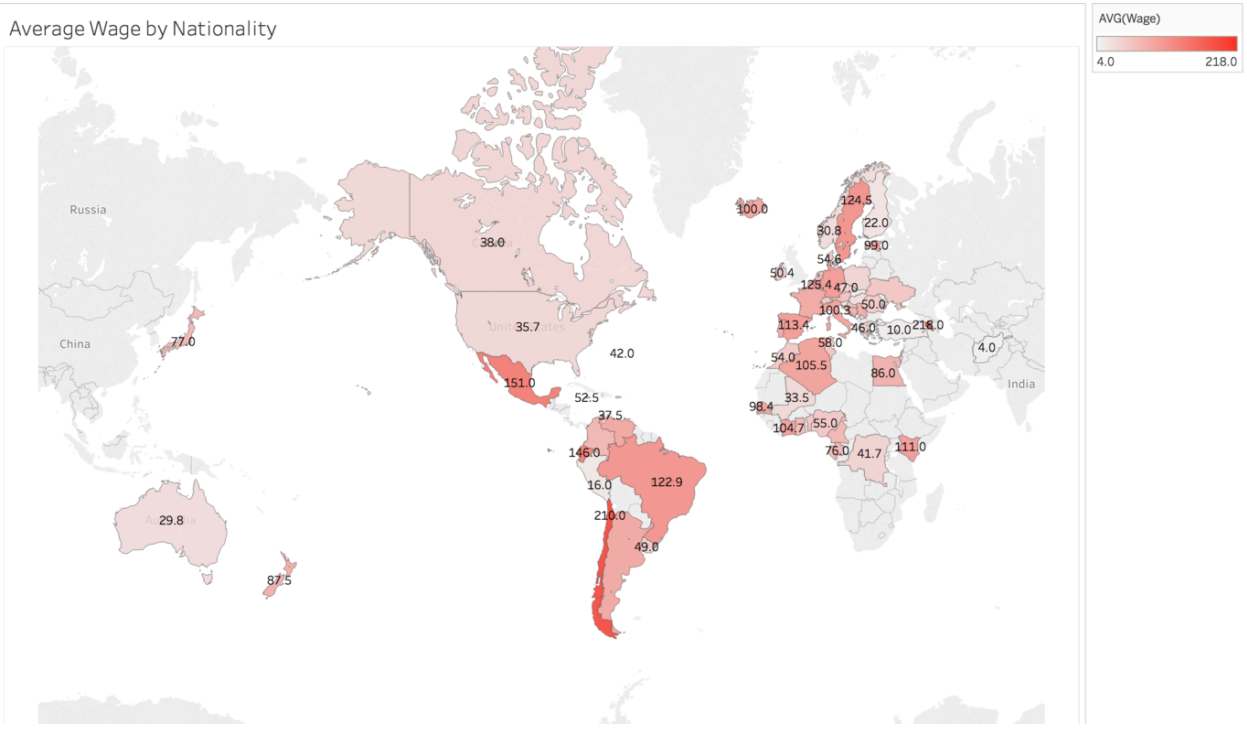

Figure 9: This figure illustrates average wage by country, coloring every country represented in the Premier League darker on the basis of higher wages.

References

2006 FA Premier League. (n.d.). (Accessed on Wed, December 06, 2017)

Arabzad, S. M., Ghorbani, M., & Shahin, A. (2013). Ranking players by DEA the case of English Premier League. *International Journal of Industrial and Systems Engineering*, *15*(4), 443-461.

Barros, C., & Leach, S. (2006). Analyzing the Performance of the English F.A. Premier League With an Econometric Frontier Model. *Journal of Sports Economics*, *7*(4), 391-407.

Campbell, D. (2002). The Observer Sport. *The Guardian.* (Accessed on Wed, December 06, 2017)

Dawson, P., Dobson, S., Goddard, J., & Wilson, J. (2007). Are football referees really biased and inconsistent?: evidence on the incidence of disciplinary sanction in the English Premier League. *Journal of the Royal Statistical Society Series A*, *170*(1), 231-250.

Dobson, S., & Goddard, J. (2010, September). Optimizing strategic behaviour in a dynamic setting in professional team sports. *European Journal of Operational Research*, *205*(3), 661-669.

Elliott, R., & Weedon, G. (2010). Foreign players in the English Premier Academy League: 'Feet-drain' or 'feet-exchange'? *International Review for the Sociology of Sport.*

Gallo, E., Grund, T., & Reade, J. J. (2012, sep). Punishing the Foreigner: Implicit Discrimination in the Premier League Based on Oppositional Identity∗. *Oxford Bulletin of Economics and Statistics*, *75*(1), 136−−156.

Goal.com. (n.d.). (Accessed on Wed, December 06, 2017)

Guanhao, F., Nicholas, P., & Jianeng, X. (2016, December). The market for English Premier League (EPL) odds. *Journal of Quantitative Analysis in Sports*, *12*(4), 167-178.

League, P. (2017). Premier League Value of Central Payments to Clubs.

Lenten, L. J., Geerling, W., & Kónya, L. (2012). A hedonic model of player wage determination from the Indian Premier League auction: Further evidence. *Sport Management Review*, *15*(1), 60-71.

Peeters, T., & Szymanski, S. (2012, Dec). *Vertical restraints in soccer: Financial fair play and the English Premier League* (Working Papers No. 2012028). University of Antwerp, Faculty of Applied Economics.

Ramasamy, R. H. . L. C. G. . B. (2002). The Socio-Economic Determinants of International Soccer Performance. *Journal of Applied Economics, Universidad del CEMA*, *5*.