English Premier League
An economics study of parameters’ impact on final position in the English Premier League 2009-2017

Simon Molin, Tom Ekelund
Abstract

Football is the most exerted sport worldwide and can arguably be considered as a major industry on its own, where the English Premier League stands out as the most popular league in the world. This thesis examines what factors that generate utility, given the assumption that clubs are utility maximization units, for the individual club in the English Premier League. Where final position is of utmost interest. For this analysis, ordered logit and multinomial logit regressions are performed through the usage of 180 observed final positions between the years of 2009-2017. Although this thesis focuses on the club’s individual utility, socio-economic utility tends to appear as a consequence when producing sports. Several parameters are discovered to have a significant and substantial impact on clubs’ final position and thus their utility.
# Table of Contents

1. Introduction ........................................................................................................... 5
2. Purpose .................................................................................................................... 5
   2.1 Formulating the problem ....................................................................................... 5
   2.2 Limitations .......................................................................................................... 6
   2.3 Problematics ........................................................................................................ 6
3. Literature review ..................................................................................................... 7
   3.1 Utility maximization function .............................................................................. 7
   3.2 Macroeconomic perspective – Competitive balance ............................................. 8
   3.3 External effects .................................................................................................... 9
4. Methodology ............................................................................................................ 10
   4.1 Discrete Choice Data .......................................................................................... 10
   4.2 Ordered logit ....................................................................................................... 10
   4.3 Multinomial logit ............................................................................................... 12
5. Data ........................................................................................................................ 13
   5.1 Turnover ............................................................................................................. 14
   5.2 NetTransfers ..................................................................................................... 15
   5.3 Squad Value ....................................................................................................... 15
   5.4 Previous ............................................................................................................. 16
   5.5 Dummy Variables .............................................................................................. 16
   5.6 MeanTemperature ............................................................................................. 16
   5.7 Competition ....................................................................................................... 17
   5.8 Income ............................................................................................................... 17
   5.9 Population .......................................................................................................... 18
   5.10 Correlation matrix ........................................................................................... 18
   5.11 Descriptive statistics ....................................................................................... 19
6. Results ...................................................................................................................... 20
   6.1 Ordered logit regression #1 ................................................................................. 20
   6.2 Ordered logit regression #2 ............................................................................... 21
   6.3 Multinomial logistic regression #1 ...................................................................... 22
   6.5 Multinomial logit regression #2 .......................................................................... 25
7. Discussion ................................................................................................................ 27
8. Conclusion ............................................................................................................... 31
9. References .............................................................................................................. 33
<table>
<thead>
<tr>
<th>Appendix</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Appendix A</td>
<td>43</td>
</tr>
<tr>
<td>Appendix B</td>
<td>43</td>
</tr>
<tr>
<td>Appendix C</td>
<td>44</td>
</tr>
</tbody>
</table>
1. Introduction

Football is one of the most exerted and beloved sports worldwide, consisting of approximately 38 million registered players combined with 270 million people directly involved in the industry (Iho & Heikkila, 2010). The English Premier League (EPL) alone, which this thesis will focus on, presents broadcasting revenues up to $13 billion between 2016-2019. The league is represented by 80 broadcasters in 212 areas around the world with an average match being watched by over 12 million people. Furthermore, comparisons of other top leagues in Europe including Spain, Italy, and Germany have been performed with the result that none of the above stated leagues even come close in comparison to EPL regarding average game viewers (Curley & Roeder, 2016). The authors noted that “The English Premier League has every right to call itself the most popular football league in the world” (p. 78). With respect to these facts, a thesis on success in the EPL would appear to be interesting since the league itself has to be considered as a major industry that seems to have an impact on both the socio-economic utility and the utility for the club itself.

2. Purpose

The purpose of this thesis is, from an economics perspective, to examine what factors that affect final position in the EPL. Furthermore, the clubs in the EPL are considered as utility maximization units where the utility perceived is mostly dependent on final position and where better position generates more utility. This is in line with Downward & Dawson’s (2000) and Sloan’s (1971) argument.

2.1 Formulating the problem

What parameters affects the EPL clubs’ utility maximization function?

\[ U = U\{P(\beta_1 + \beta_2 + \beta_k), X, (\pi_R - \pi_0 - T)\} \]

The above stated equation describes a variant of Sloan’s (1971) equation where each club’s utility perceived depends on \( P \) = final position in the EPL which itself is affected by \( \beta \). To be able to experience utility, the clubs need to face other clubs, here an assumption is taken; the more competitive the league is \( (X) \), the more utility each club will experience. The perceived
utility is subject to a financial constraint where \( \pi_R \) is clubs’ profit, \( \pi_0 \) is minimum acceptable profit and \( T \) is tax. The main objective in this thesis is to examine the effect of \( \beta \).

### 2.2 Limitations

Portsmouth encountered financial issues during the 2009/2010 season. In November 2009 they were not capable of paying salary to their players and staff due to administration failures and unsuccessful player transfers. This led them to the verge of bankruptcy. As a result of these financial problems, Portsmouth FC fails to provide published financial reports for the 2008/2009 and 2009/2010 seasons. Therefore, an estimation of Portsmouth’s financial variables covering Turnover, Net Transfers and Squad Value are calculated in a separate manner. See explanation of the calculation in appendix A.

### 2.3 Problematics

In previous research, home league match attendance has been used as a predictor to determine final position in the league (Pinnuck & Potter, 2006). In this thesis, turnover consist of home league match revenue and is therefore a representation of home league match attendance. However, home league match attendance in the EPL could arguably be misleading since the price elasticity of attending a league game is very low. (Dobson & Goddard, 1995; Szymanski & Kuypers, 1999). Since all teams in the EPL more or less sell out their stadiums, home league match attendance is dependent on the capacity of the stadiums and do not serve its real cause since demand is higher than supply. Therefore, a separate variable for home league match attendance will not be included.
3. Literature review

Several sources have contributed to this thesis, mostly consisting of journals and books. Yet, Downward & Dawsons’ (2000) book ‘The Economics of Professional Team Sports’ has provided the fundamental base of this paper.

3.1 Utility maximization function

Downward & Dawson (2000) explicitly explain the economics connection to the sports industry. From a micro economic perspective, labor can be described as players and coaching staff. Simultaneously, capital can be described as stadiums which together creates the final product; sport. This product is then sold to consumers which in this instance consist of fans and spectators. The authors make a distinction between the American and the European sport industry. The objective for American professional sports teams is to maximize profit since they operate in a more commercial context with closed leagues (teams can’t be relegated). Furthermore, they argue that the objectives for English sports teams are to maximize the utility. The reasons behind this assumption is that owners’ investments in clubs are made to give them satisfaction and not profit. Owners of clubs and the club itself tend to lose money over long periods, but they still continue to operate despite the fact that by leaving the sports industry losses could be prohibited. Sloan, (1971) also argues that the objective for football clubs is to maximize utility and that the objective criteria for maximizing the utility are relatively clear, to have playing success. The goal for EPL clubs is therefore to maximize utility, subject to financial constraints, through sporting success which is measured by the final position in the league. Sloan, (1971) display the following utility maximization function as:

\[ U = U(A, P, X, (\pi_R - \pi_0 - T)) \]

Where \( U \) = Utility perceived by clubs depends on, \( A = \) average home match attendance, \( P = \) on field performance and \( X = \) competitive balance. Utility is subject to a financial constraint where \( \pi_R \) is clubs’ profit, \( \pi_0 \) is minimum acceptable profit and \( T \) is tax. The main difference between Sloan’s (1971) utility function and the function stated in the ‘formulating the problem’ section is that \( A \) (home league match attendance) will estimate final position through the independent variable Turnover (see explanation in the ‘Problematics’ section).
3.2 Macroeconomic perspective – Competitive balance

The fact that clubs in the English Premier League are competing against other clubs guarantees a competitive balance which is essential in order to generate utility for the clubs. Prior research has argued that increased competitiveness and, more importantly, increased uncertainty over the outcome will produce higher interest and support from fans (Rottenberg, 1956), (Forres & Simmons, 2002), (Pinnuck & Potter, 2006). However, the EPL is becoming more and more unbalanced with similar teams finishing in the top positions, which endangers the “uncertainty of outcome” principle (Salomon Brothers, 1997). A macro-economic perspective on this phenomenon is taken by Curley & Roeder (2016) who demonstrates the competitive balance in European Football Leagues by partly using the Gini-coefficient. The leagues observed are the EPL, La Liga, Serie A, Bundesliga, Ligue 1 and Eredivisie. Moreover, the Gini-coefficient is commonly used to describe inequality, and in this case how points are allocated. Their results show that the EPL has consistently been the most unbalanced league, where the distribution of points won are being allocated to fewer teams. Because of this, it has been claimed that regulations regarding the allocation of teams’ resources should be regulated by league authorities. This, in order to guarantee a competitive balance that will ensure the success and survival of the league (Downward & Dawson, 2000). Regulations have been implemented to ensure a competitive balance in European club football. In 2009 UEFA approved a Financial Fair Play concept to protect the game and its well-being. UEFA states that the main objectives of the Financial Fair Play are to encourage clubs to operate on the basis of their own revenue sources, to encourage responsible spending in transfers of players and to improve the overall financial capabilities of clubs (UEFA, 2019).

Sporting success in the EPL is arguably linked to the financial performance of a football club. Clubs who are the market leaders, in terms of turnover, appear to have a much higher chance of success. Clubs with lower turnover appears to play sub-championships of their own and have different objectives (Barros & Leach, 2006). This is logical since there are teams who seem to be satisfied with simply staying in the EPL and do not aspire to achieve the top positions. Since every club’s utility maximization function is subject to a financial constrain, it could be argued that they still maximize their utility given their individual objectives. Supporting these statements are various strategic reports in their closings e.g., Burnley FC and Watford FC among many (Beta Companies House, 2019).
3.3 External effects

A recurrent phenomenon in economic theory is the notion regarding external effects, perhaps most commonly known from environmental concerns related to negative externalities with Arthur Pigou as a front figure. Should for example, the benefits or the costs for a transaction transcend to a third party, externalities are considered as present. Social benefits could be revenues generated by the media that derives from a club’s activities. A market failure is defined when only private costs and benefits are accounted for, and not social costs and benefits. Therefore, it could be argued that sports are under produced as a commodity due to negligence of social costs and benefits (Downward & Dawson, 2000). Furthermore, as the commodity could be considered as price inelastic it also supports the lack of production since demand is substantially higher than supply in the EPL (Dobson & Goddard, 1995; Szymanski & Kuypers, 1999). Here, authors Gouguet and Barget (as cited in Andreff & Szymanski, 2006) demonstrate a connection to sports where externalities emerge in both positive and negative fashions. Gouguet and Barget continue their reasoning by dividing sport externalities into two categories; Social links and territorial dynamics. Where positive social links refer to social cohesion (local patriotism) and social recognition, especially in terms of ethnic minorities and people from disadvantaged areas. Simultaneously, positive territorial dynamics could appear in terms of lessening of tensions linked to unemployment, misdemeanor and drugs. Moreover, this might result in a strengthen brand image which improves the attractiveness of the club even further. Arguably, investors’ interest may arise and the individual club’s economy might blossom.

On the negative side however, social cohesion may occur in forms of hooliganism along with cheating, doping and other underground activities that suffer credibility. Ultimately, individuals jeopardize their chances to ever participate and attend a sporting event. Gouguet and Barget (as cited in Andreff & Szymanski, 2006), also states that negative territorial dynamics appear in terms of crowding-out effects, where individuals might experience saturation due to major sport events. It could be argued that circumstances like these fend off external investors since association with clubs that have a poor reputation is not desirable.

Yet, this thesis interprets the clubs as utility maximization units that produce the final product which is sport, as Sloan (1971) have demonstrated. Externalities due to sports will be further evaluated in the discussion part.
4. Methodology

Two regression methods will be used to examine what factors that influence clubs’ utility functions. The regressions are estimated in the program STATA.

4.1 Discrete Choice Data

A discrete choice or multiple-choice variable indicates that the variable can adopt numerous unordered or ordered qualitative values (Stock & Watson, 2014). Models that employ discrete choice data chooses between two or more discrete alternatives to try and explain and predict the outcome. In comparison to consumption theory that utilizes continuous data with focus on “how much”, discrete choice data emphasizes on “which one” (Train, 1986). In this analysis, the dependent variable is final position in the EPL. A club may finish in the top, middle or bottom of the table. Hence, the dependent variable has three possible outcomes where these outcomes are ordered. Therefore, these outcomes depend on certain qualitative characteristics (Stock & Watson, 2014). In this analysis, these characteristics are e.g., turnover, squad value and position last season among others. Moreover, Stock & Watson (2014) argue that the econometric task is to model the probability of ending up in one of the possible outcomes, given the individual characteristics. Yet, the authors recommend that models concerning the analysis of discrete choice data are feasible to interpret via principles of utility maximization, which this thesis revolves around (higher position equals higher utility). Whereas this is best performed through usage of probit or logit regression models. This thesis will take two different methods concerning discrete choice data into consideration; Ordered logit regression and Multinomial logit regression.

4.2 Ordered logit

Viewed as an extension of a logistic regression model, an ordered logit can prove useful when predicting a response originating from different characteristics within the independent variables (McCullagh, 1980). Gray, Rivero-Arias & Clarke (2006) supports this choice of model by deciding whether the dependent variables are categorical with discrete outcomes or not. If they are, a legit option would be to perform an ordered logit to predict the probability of each response level. Related to this thesis, this would be the three levels Top (3), Middle (2) or Bottom (1) of the EPL league table. Other arguments promoting an ordered logit is the fact that it recognizes that each categorical response is ordered (Zavoina & Mc Kelvey, 1975). However,
for an ordered logit to prove useful and valid, it needs a certain criterion to be fulfilled; the Proportional Odds Assumption. It essentially means that the predictors have the same effect on the odds of moving to a higher-order category everywhere along the scale. Thus, parallel slopes characterize the cumulative probability curves for each of the ordered categories. If the Proportional Odds Assumption is violated, a multinomial logit could prove useful since it provides parameter estimates that are unbiased (Nerlove & Press, 1973).

In this thesis, two tests are performed to check whether the Proportional Odds Assumption is violated. Firstly, a likelihood ratio test that originates from a user-written command by Rory Wolfe & William Gould (1999). Here, the null hypothesis indicates no difference in the coefficients between models, hence, a non-significant result is desirable. Secondly, a Brant test is performed which compares slope coefficients of the k-1 binary logits implied by the ordered regression model. A non-significant result is desirable even here (Long & Freese, 2013).

The equation for an ordered logit regression model is usually displayed as:

\[
\text{logit}(p_1 + p_2 + \ldots + p_k) \equiv \log \frac{p_1 + p_2 + \ldots + p_k}{1 - p_1 + p_2 + \ldots + p_k} = \alpha_k + \beta x
\]

where \( p_1 + p_2 + \ldots + p_k = 1 \)

Ordered logit models have a cumulative probability. It simultaneously estimates multiple equations where it only estimates the number of categories in the dependent variable minus one (k-1). In this instance, two equations will be estimated since we have three categories. These would be displayed as:

\[
\text{logit}(p_3) \equiv \log \frac{p_3}{1 - p_3} = \alpha_3 + \beta x
\]

\[
\text{logit}(p_3 + p_2) \equiv \log \frac{p_3 + p_2}{1 - p_3 - p_2} = \alpha_2 + \beta x
\]
4.3 Multinomial logit

In contrast to the ordered logit, the multinomial logit assumes no order to the categories of the dependent variable since the categories are nominal (UCLA: Statistical Consulting Group, 2018), moreover, as stated above, it can produce unbiased results if the Proportional Odds Assumption from the ordered logit is violated (Nerlove & Press, 1973). However, conclusions of the result should be taken with caution since one may not be answering the research question one is really interested in, if it incorporates the ordering (Grace-Martin, 2019) Multinomial probit and logit regressions tend to produce similar results, hence, the choice of method is not of vital importance (Stock & Watson, 2014). In this thesis, the analysis will be based on a multinomial logit regression model. By usage of multinomial logit, the dependent variable may adopt the values 1, 2 or 3, where the model tries to predict group affiliation given the independent variables. Furthermore, a value of 3 indicates a top position ranging from 1 to 4. These positions are of particular interest in the EPL since it translates into Champions League participation the following year. Clubs receiving a position between 18 to 20 will adopt the value 1, which indicates a bottom position and immediate relegation. Remaining positions, 5 to 17 will adopt the value 2 which indicates a middle position. Moreover, the middle position will be used as the reference category in this instance since predictors that affects top tier and bottom tier positions is of greater interest. Grace-Martin (2019) states that every predictor’s effect tells the probability of success in a certain category (Top or Bottom) in comparison to the reference category, the middle. Ultimately, each category has its own coefficients and intercept which means that every predictor can affect each category differently. Therefore, the probability of success in terms of Top and Bottom may differ even though it is the same predictor.

The equation for a multinomial regression model is usually displayed as:

$$Prob(Y_t = j|w_t) = P_{ij} = \frac{\exp(w'_t a_j)}{1 + \sum_{k=1}^{J} \exp(w'_t a_k)}$$

where $j = 1,2...k$ $Y$ is the observed outcome and exp is the base of the natural logarithm. $w'$ is a vector for a predictor and $a$ is a parameter (Greene, 2008).
5. Data

The data contains 180 observations of the dependent variable, final position. There are 34 different teams in the EPL between the years 2009-2017. There are eight teams who have participated in all of the consecutive 9 seasons in the EPL, these are; Arsenal, Chelsea, Liverpool, Manchester City, Manchester United, Stoke, Sunderland and Tottenham. The remaining teams have played at least one or more seasons in one of the leagues below the EPL.

Figure 1
Table 1, Descriptive summary of independent variables used to predict final position.

<table>
<thead>
<tr>
<th>Variable name</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Turnover</td>
<td>Clubs’ seasonal turnover expressed in tenth of millions £</td>
</tr>
<tr>
<td>Competition</td>
<td>Competition in cities regarding other teams in the EPL</td>
</tr>
<tr>
<td>Previous</td>
<td>Previous year’s final position</td>
</tr>
<tr>
<td>MeanTemperature</td>
<td>Mean temperature expressed in Celsius</td>
</tr>
<tr>
<td>Income</td>
<td>Regional gross disposable household income per head expressed in thousands of £</td>
</tr>
<tr>
<td>Population</td>
<td>Major town and city population expressed in tenth of thousands £</td>
</tr>
<tr>
<td>Net Transfers</td>
<td>Additions subtracted with disposal of players expressed in millions of £</td>
</tr>
<tr>
<td>Squad Value</td>
<td>Total market value of all players expressed in millions of £</td>
</tr>
</tbody>
</table>

In an alternative model, Dummy variables will be used for previous year’s final position and replace the independent variable Previous. The reason to include Dummy variables instead of one independent variable for previous year’s final position is to isolate different effects, depending whether the club finished top, middle, bottom or was promoted the previous year.

5.1 Turnover

The independent variable Turnover consist of matchday revenues, commercial and sponsorship incomes and broadcasting revenues on a seasonal basis and are expressed in tenth of millions £. Notably, transfer fees and revenues from disposals of players are not included in Turnover. Hence, two separate independent variables, Squad value and Net transfers, are used to take this
effect into account. For most of the clubs in the EPL broadcasting revenues has the biggest share of total turnover. There are a few clubs who have large commercial and sponsorship incomes that exceeds broadcasting revenues. The hypothesis for this variable is that it will have a positive impact on the dependent variable, where clubs with high turnover will obtain a high position in the league. This hypothesis is in line with previous literature who has concluded that revenue is positively related to playing success (Pinnuck & Potter, 2006), (Barros & Leach, 2006), (Carmichael, McHale & Thomas, 2010), (El-Hodiri & Quirk, 1971), (Szymanski & Kuypers, 1999). All teams’ turnover has been obtained via each club’s individual annual financial report.

5.2 NetTransfers

In clubs’ annual financial report, the cost associated with acquisition of players are capitalized as intangible fixed assets. NetTransfers have been calculated by taking the additions subtracted with disposals of intangible fixed assets. Where a positive number indicates that additions exceed disposals in terms of player trading. Hence, the total squad value increases. NetTransfers is an indicator on how active a club has been on the transfer market. An increase in total squad value would, arguably, lead to a better position than last season (Carmichael, McHale & Thomas, 2010). Observations in clubs’ annual financial reports has shown an overall increase in total squad value between the observed years 2009-2017. This concludes that acquisitions of players surpass disposals. Therefore, the hypothesis with regards to this independent variable is that it will affect final position positively.

5.3 Squad Value

The squad value is the total value of intangible fixed assets in millions of £, also collected from club’s annual financial reports. Where the net transfers of players have been added to the historical total value of intangible fixed assets. This independent variable will try to reflect the players’ capacity and level of skill on the pitch. Naturally, more skilled and competent players should give the club better possibilities to finish on a top position. Hence, the hypothesis for the independent variable Squad Value is that it will have a positive effect on the final position in the league.
5.4 Previous

Previous is an independent variable which represents teams’ final position from the previous year in the EPL. In the regressions, teams will be indicated with number 1 if they were placed in the top four teams of the league, number 3 if they were placed in the bottom three teams, 15-17th, since relegated teams are not included in Previous. Number 2 represents the rest of the teams in the EPL the previous season. Teams who were being promoted to the EPL from the last season will be indicated with number 4. The hypothesis for this variable is that it will have a positive effect on the final position, due to observed observations of league tables, where teams who end up with a top position the last year usually obtain a top position the following year. Furthermore, expectations regarding low fluctuations of the remaining teams also strengthen this hypothesis.

5.5 Dummy Variables

The creation of Dummy variables is made in order to replace the independent variable Previous and have been separated into four different categories, since it is believed to provide a more isolated effect on each category. These variables are: DummyTop, DummyMid, DummyBottom and DummyPromotion. Each different Dummy variable will be designated with number 1 for specific positions in the previous year’s final position and number 0 if they did not obtain that position. For example, DummyTop will be indicated with number 1 for the four top teams the previous position and 0 for all other positions. The hypothesis for these variables is that top teams the previous year will have a positive effect on the dependent variable. Furthermore, teams in the bottom and promoted teams from last year will remain in the bottom or be relegated and therefore have a negative effect on the dependent variable.

5.6 MeanTemperature

Mean Temperature is obtained at the closest meteorological station from the city where the club is based and is measured in average temperature for one year in degrees of Celsius. The hypothesis presented here is intuitive, warmer mean temperature may provide better prerequisites and thus effect final position positively. Welki & Zlatoper (1999) found that warmer temperature had a positive impact on the demand for football in Sweden where the mean temperature differs between regions. However, the mean temperature is similar across
the UK and the expectations on the influence on final position is low. Furthermore, mean temperature might have influenced the sporting culture historically, both on a global and regional basis within a country. Which could have an impact on today's interest and popularity of clubs in different regions.

5.7 Competition

The interpretation of competition is that teams in the EPL that are playing in the same city will have a higher degree of competitiveness. In this thesis, the degree of competitiveness is ranging between 0-5. Where 5 indicates that it resides six teams in a city (this is the highest number of teams that simultaneously have participated in the EPL from the same city during the observed years), thus, competition is regarded as high. Furthermore, 0 indicates no competition at all. One exception that deviates from this criterion are the two clubs represented from Wales; Cardiff and Swansea. Two different cities but are considered rivals according to the public (James, 2013). Previous research suggest that higher competition arouse spectator interest which ultimately leads to higher demand and commercial activity for each club (Downward & Dawson, 2000). It could be argued that a higher degree of competition thus leads to better performance as a consequence of increased demand, turnover etc. The hypothesis regarding this variable is therefore concluded as greater competitiveness equals greater final position.

5.8 Income

Data on Income is gathered from the Office of National Statistics and expressed in thousands of £ at 2016 basic prices. Income is defined as regional gross disposable household income and is the amount of money that all individuals in the household have available for spending or saving after taxes and any direct benefits. It is also measured per head, which is estimates of values for each person (Office for National Statistics, 2019b). Arguably, better income per head could be reflected as better economical wealth for regions, and therefore better prerequisites. Because of this, the plausibility of investments in local sports might increase. Therefore, the hypothesis is that higher Income affects final position positively. Furthermore, Leeds & Leeds (2009) supports this hypothesis by stating that positive effects of BNP/capita can increase sporting success.
The data on income from the Office of National Statistics is lagged two years and is therefore not available for 2017. Thus, estimated calculations for regional gross disposable household income per head for 2017 are produced and can be found in Appendix B.

5.9 Population

The independent variable Population contains major town and city population expressed in tenth of thousands, retrieved from the Office for National Statistics. Moreover, the estimated resident population considers the following: arriving international migrants if they remain in the UK for at least a year. Simultaneously, emigrants are excluded if they remain outside the UK for at least a year (Office for National Statistics, 2019a). The hypothesis for this explanatory variable is that larger population affects final position positively. This hypothesis is based on El-Hodiri & Quirk (1971) and Szymanski & Kuypers (1999) studies who claims that larger populated cities have better absorption capabilities of players and supporters. Both better playing performance and greater match attendance generates turnover. Furthermore, the authors claim that more revenue is connected to a better long-term performance.

5.10 Correlation matrix

Finally, a correlation matrix is performed to measure whether the predictors are associated with each other. Where a correlation coefficient exceeding 0.7 could be considered as strong correlation (Schober, Boer & Schwarte, 2018). As VIF-tests cannot be performed through STATA when running ordered logits (Williams, 2011) a common rule of thumb, according to Dr. Baral (Researchgate, 2014) is to acknowledge multicollinearity when the correlation coefficient exceeds 0.8.

Table 2, Correlation Matrix

<table>
<thead>
<tr>
<th>Correlation Matrix</th>
<th>Turnover</th>
<th>Competition</th>
<th>DummyTop</th>
<th>MeanTemperature</th>
<th>Income</th>
<th>Population</th>
<th>SquadValue</th>
</tr>
</thead>
<tbody>
<tr>
<td>Turnover</td>
<td>1,000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Competition</td>
<td>0.250</td>
<td>1,000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DummyTop</td>
<td><strong>0.725</strong></td>
<td>0.270</td>
<td>1,000</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MeanTemperature</td>
<td>0.299</td>
<td><strong>0.834</strong></td>
<td>0.318</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Income</td>
<td>0.282</td>
<td><strong>0.830</strong></td>
<td>0.219</td>
<td><strong>0.877</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Population</td>
<td>0.227</td>
<td><strong>0.952</strong></td>
<td>0.272</td>
<td><strong>0.909</strong></td>
<td><strong>0.899</strong></td>
<td>1,000</td>
<td></td>
</tr>
<tr>
<td>SquadValue</td>
<td><strong>0.884</strong></td>
<td>0.224</td>
<td>0.654</td>
<td>0.273</td>
<td>0.226</td>
<td><strong>0.187</strong></td>
<td>1,000</td>
</tr>
</tbody>
</table>
5.11 Descriptive statistics

Table 3 displays descriptive statistics of notable independent variables. Notably, Turnover, NetTransfers, Population and SquadValue presents large variation in min and max values. Where SquadValue displays the largest variation, with a min value of 5,8 and max value of 687,9 in millions of £.

Table 3

<table>
<thead>
<tr>
<th>Independent Variable</th>
<th>Mean</th>
<th>Std. Dev.</th>
<th>Min</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>Turnover</td>
<td>14,6</td>
<td>10,8</td>
<td>2,2</td>
<td>52,1</td>
</tr>
<tr>
<td>NetTransfers</td>
<td>19,7</td>
<td>35,6</td>
<td>-74</td>
<td>167,7</td>
</tr>
<tr>
<td>Income</td>
<td>18,4</td>
<td>4,4</td>
<td>13,6</td>
<td>28</td>
</tr>
<tr>
<td>Population</td>
<td>270,2</td>
<td>360,8</td>
<td>8,2</td>
<td>878,9</td>
</tr>
<tr>
<td>SquadValue</td>
<td>142,4</td>
<td>129,4</td>
<td>5,8</td>
<td>687,9</td>
</tr>
</tbody>
</table>

Table 4 presents the club’s allocation regarding final position over the nine observed years. Where ‘Position 3’ indicates a top four position, ‘2’ indicates a middle position ranging from 5-17 and ‘1’ indicates a bottom three position.

Table 4

<table>
<thead>
<tr>
<th>Position</th>
<th>Freq.</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>27</td>
</tr>
<tr>
<td>2</td>
<td>117</td>
</tr>
<tr>
<td>3</td>
<td>36</td>
</tr>
</tbody>
</table>
6. Results

The format used to interpret the coefficients for all of the following ordered logit regressions are established through proportional odds ratios. The sole purpose being easier interpretations of the coefficients. As previously stated in the methodology section, two tests are performed for all ordered logit regressions to check whether the Proportional Odds Assumption is violated; A likelihood ratio test and a Brant test.

6.1 Ordered logit regression #1

Table 5, Ordered logit regression #1

<table>
<thead>
<tr>
<th>Independent Variable</th>
<th>β</th>
<th>Robust S.E</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Turnover</td>
<td>1,082</td>
<td>0,060</td>
<td>0,160</td>
</tr>
<tr>
<td>Competition</td>
<td>0,773</td>
<td>0,245</td>
<td>0,419</td>
</tr>
<tr>
<td>Previous</td>
<td>0,507**</td>
<td>0,119</td>
<td>0,004</td>
</tr>
<tr>
<td>MeanTemperature</td>
<td>2,523**</td>
<td>0,952</td>
<td>0,014</td>
</tr>
<tr>
<td>Income</td>
<td>0,775**</td>
<td>0,091</td>
<td>0,032</td>
</tr>
<tr>
<td>Population</td>
<td>1,002</td>
<td>0,002</td>
<td>0,328</td>
</tr>
<tr>
<td>NetTransfers</td>
<td>0,982**</td>
<td>0,006</td>
<td>0,011</td>
</tr>
<tr>
<td>SquadValue</td>
<td>1,008*</td>
<td>0,005</td>
<td>0,075</td>
</tr>
</tbody>
</table>

Proportional Odds Assumption

<table>
<thead>
<tr>
<th>Test</th>
<th>p&gt;chi2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Likelihood Ratio Test</td>
<td>0,124</td>
</tr>
<tr>
<td>Brant Test</td>
<td>0,115</td>
</tr>
</tbody>
</table>

* Significance level 10 %    ** Significance level 5 %    *** Significance level 1 %

The first regression performed is a random-effects ordered logistic regression including all independent variables from above stated data, except Dummy Variables since Previous is included. The likelihood ratio test presents a prob>chi2 value of 0,1240 and the Brant test displays a p>chi2 value of 0,115. This indicates a non-significant result; thus, the Proportional Odds Assumption is not violated and allows for further analysis (SAS Institute inc., 2019). In accordance with Huber, clustered robust standard errors are applied that allow for additional correlation within panels (StataCorp LLC, 2013).
Notable mentions for significant\(^1\) independent variables, given that all of the other variables are held constant, in regression #1:

For one unit increase in Previous, i.e., going from bottom to middle the previous year, the odds of a top position versus the combined middle and bottom position are 0.507 times greater.\(^2\) For one-degree Celsius increase in MeanTemperature the odds of observing a top position is 2.523 times greater than observing a combined middle and bottom position. For the predictor Income, one-unit increase (thousand £), the odds of a top position versus the combined middle and bottom positions are 0.775 times greater. Furthermore, for one-unit increase (one million £) in NetTransfers, the odds of a top position versus the combined middle and bottom positions are 0.982 times greater. While for one-unit increase (1 million £) in SquadValue, the odds of a top position versus the combined middle and bottom positions are 1.008 times greater.

### 6.2 Ordered logit regression #2

In this ordered logit regression, the predictor Previous has been omitted in favor of Dummy Variables. DummyMiddle has been excluded since a middle position is arguably the least interesting since bottom and top teams are affected more significant of their final position. This, in terms of relegation and Champions League qualifications the following year.

However, interpretations of the odds ratios in the ordered logit regression #2 should be taken with caution since the Proportional Odds Assumption is violated. The likelihood ratio test displays a p>chi2 value of 0.000 which concludes that the predictors’ odds of moving to a higher-order category everywhere along the scale, have different effects. For interpretations regarding ordered logit regression #2 see appendix C.

According to Grace-Martin (2019) and Nerlove & Press (1973), results should be reported in multinomial logit if the Proportional Odds Assumption is violated since it provides unbiased parameter estimates. Multinomial logit regressions are therefore performed.

---

\(^1\) Significant levels range between 1 to 10 percent. See table 5 for each variable’s specific significant level.  
\(^2\) Note that odds ratios lower than one affects final position negatively.
6.3 Multinomial logistic regression #1

The format used to interpret the coefficients for the all of following multinomial logistic regressions are established through relative risk ratios. The sole purpose being easier interpretations of the coefficients. Moreover, clustered robust standard errors are applied that allow for additional correlation within panels (StataCorp LLC, 2013).

Table 6, Multinomial logit regression #1

<table>
<thead>
<tr>
<th>Independent Variable</th>
<th>( \beta )</th>
<th>Robust S.E</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Turnover</td>
<td>1,089</td>
<td>1,064</td>
<td>0,149</td>
</tr>
<tr>
<td>Competition</td>
<td>0,850</td>
<td>0,661</td>
<td>0,835</td>
</tr>
<tr>
<td>DummyTop</td>
<td>2,280</td>
<td>1,737</td>
<td>0,279</td>
</tr>
<tr>
<td>DummyBottom</td>
<td>0,000***</td>
<td>0,000</td>
<td>0,000</td>
</tr>
<tr>
<td>DummyPromotion</td>
<td>0,000***</td>
<td>0,000</td>
<td>0,000</td>
</tr>
<tr>
<td>MeanTemperature</td>
<td>2,831</td>
<td>3,792</td>
<td>0,437</td>
</tr>
<tr>
<td>Income</td>
<td>0,711</td>
<td>0,180</td>
<td>0,181</td>
</tr>
<tr>
<td>Population</td>
<td>1,002</td>
<td>0,005</td>
<td>0,623</td>
</tr>
<tr>
<td>NetTransfers</td>
<td>0,970***</td>
<td>0,008</td>
<td>0,001</td>
</tr>
<tr>
<td>SquadValue</td>
<td>1,010*</td>
<td>0,005</td>
<td>0,061</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Independent Variable</th>
<th>( \beta )</th>
<th>Robust S.E</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Turnover</td>
<td>0,838</td>
<td>0,102</td>
<td>0,151</td>
</tr>
<tr>
<td>Competition</td>
<td>1,101</td>
<td>0,474</td>
<td>0,823</td>
</tr>
<tr>
<td>DummyTop</td>
<td>0,000***</td>
<td>0,000</td>
<td>0,000</td>
</tr>
<tr>
<td>DummyBottom</td>
<td>5,419***</td>
<td>3,260</td>
<td>0,005</td>
</tr>
<tr>
<td>DummyPromotion</td>
<td>3,952*</td>
<td>2,804</td>
<td>0,053</td>
</tr>
<tr>
<td>MeanTemperature</td>
<td>0,407</td>
<td>0,223</td>
<td>0,102</td>
</tr>
<tr>
<td>Income</td>
<td>1,458**</td>
<td>0,251</td>
<td>0,029</td>
</tr>
<tr>
<td>Population</td>
<td>0,998</td>
<td>0,003</td>
<td>0,661</td>
</tr>
<tr>
<td>NetTransfers</td>
<td>0,983</td>
<td>0,016</td>
<td>0,322</td>
</tr>
<tr>
<td>SquadValue</td>
<td>3,008</td>
<td>11,812</td>
<td>0,779</td>
</tr>
</tbody>
</table>

* Significance level 10 %   ** Significance level 5 %   *** Significance level 1 %
Notable mentions for independent variables, given that all of the other variables are held constant, in multinomial logit regression #1;

Given a unit increase (ten million £) in Turnover, the relative risk of being in the top positions are $1,089^3$ times more likely than being in the middle positions. Furthermore, the relative risk of being in the bottom positions, given a ten million £ increase in Turnover, are $0.838^4$ times more likely than being in the middle positions. Note that Turnover has a p-value of 0.149 for Top and 0.151 for bottom and is only significant on a 20% level. This is in line with previous ordered logit regressions, where increased Turnover indicates better final position.

Regarding the predictor DummyTop, the relative risk of being in the top positions are 2,280 times more likely than being in the middle positions, however insignificant. Moreover, the relative risk of being in the bottom positions are 0.000 times more likely than obtaining a middle position. The predictor DummyBottom is significant for both top and bottom positions. The relative risk of being in the top positions are 0.000 times more likely than being in the middle positions. Analogously, the relative risk of being in the bottom positions are 5,419 times more likely than being in the reference category. As for the last dummy, DummyPromotion, the relative risk of being in the top positions are 0,000 times more likely than obtaining a middle position. Simultaneously, the relative risk of being in the bottom positions are 3,952 times more likely than being in the middle positions. Where this predictor is significant for both top and bottom positions. These results, regarding the dummy variables, are in accordance with the ordered logit regressions.

Given a unit increase (one thousand £) in Income, the relative risk of being in the top positions are 0,711 times more likely than being in the middle category, but insignificant. Yet, for bottom positions Income is significant and the relative risk of being in the bottom positions are 1,458 times more likely than obtaining a middle position. These results are analogously with the ordered logit regressions.

---

3 RRR-values over 1 indicates a higher probability of being in the comparison group (Top or Bottom).
4 RRR-values under 1 indicates a higher probability of being in the reference group (Middle).
For one-unit increase (one million £) in NetTransfers, the relative risk of being in the top positions are 0.970 times more likely than being in the middle positions. The relative risk of obtaining a bottom position is 0.9983 times more likely than obtaining a middle position. Note that it is insignificant for bottom positions. For the ordered logit regression NetTransfers affected final position negatively and which goes in line with top positions but not bottom positions for the multinomial logit regression.

Finally, given one-unit increase (one million £) in SquadValue, the relative risk of being in the top positions are 1.010 times more likely than being in the reference category. On the contrary, the p-value regarding SquadValue for bottom positions is very insignificant (0.779) and will therefore not be interpreted. The interpretation of SquadValue is in accordance with the ordered logit regression.

Notably in the correlation matrix (see table 2), Competition, MeanTemperature, Income and Population all correlates with each other over 0.8, which is an indication of high correlation (Schober, Boer & Schwarte, 2018). Therefore, Competition and MeanTemperature will be omitted in the next multinomial logit regression. Moreover, SquadValue will also be omitted as a result of high correlation with Turnover in combination with high p-values.
### 6.5 Multinomial logit regression #2

Table 7, Multinomial logit regression #2

<table>
<thead>
<tr>
<th>Top positions</th>
<th>Independent Variable</th>
<th>$\beta$</th>
<th>Robust S.E</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Turnover</td>
<td>1,191***</td>
<td>0,065</td>
<td>0,002</td>
<td></td>
</tr>
<tr>
<td>DummyTop</td>
<td>2,440</td>
<td>1,707</td>
<td>0,202</td>
<td></td>
</tr>
<tr>
<td>DummyBottom</td>
<td>0,000***</td>
<td>0,000</td>
<td>0,000</td>
<td></td>
</tr>
<tr>
<td>DummyPromotion</td>
<td>0,000***</td>
<td>0,000</td>
<td>0,000</td>
<td></td>
</tr>
<tr>
<td>Income</td>
<td>0,690**</td>
<td>0,120</td>
<td>0,033</td>
<td></td>
</tr>
<tr>
<td>Population</td>
<td>1,005**</td>
<td>0,002</td>
<td>0,021</td>
<td></td>
</tr>
<tr>
<td>NetTransfers</td>
<td>0,975**</td>
<td>0,009</td>
<td>0,013</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Bottom positions</th>
<th>Independent Variable</th>
<th>$\beta$</th>
<th>Robust S.E</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Turnover</td>
<td>0,845***</td>
<td>0,048</td>
<td>0,003</td>
<td></td>
</tr>
<tr>
<td>DummyTop</td>
<td>0,000***</td>
<td>0,000</td>
<td>0,000</td>
<td></td>
</tr>
<tr>
<td>DummyBottom</td>
<td>6,363***</td>
<td>3,441</td>
<td>0,001</td>
<td></td>
</tr>
<tr>
<td>DummyPromotion</td>
<td>5,800***</td>
<td>3,181</td>
<td>0,001</td>
<td></td>
</tr>
<tr>
<td>Income</td>
<td>1,249</td>
<td>0,189</td>
<td>0,140</td>
<td></td>
</tr>
<tr>
<td>Population</td>
<td>0,997</td>
<td>0,001</td>
<td>0,166</td>
<td></td>
</tr>
<tr>
<td>NetTransfers</td>
<td>0,980</td>
<td>0,015</td>
<td>0,208</td>
<td></td>
</tr>
</tbody>
</table>

* Significance level 10 %    ** Significance level 5 %    *** Significance level 1 %

Notable mentions for independent variables, given that all of the other variables are held constant, in multinomial logit regression #2:

Turnover is now significant for both top and bottom positions. The relative risk ratio of being in the top positions are 1,191 times more likely than the middle positions. For the bottom positions the relative risk ratio are 0,845 times more likely than the reference category.
Respective figures for DummyBottom and DummyPromotion are intact from multinomial logit regression #1 regarding top positions. For bottom positions however, the relative risk ratio has increased to 6,363 times for DummyBottom and to 5,800 times for DummyPromotion with higher significance.

Given a one-unit increase (thousand £) in Income, the relative risk of being in the top positions are now 0.690 times more likely than being in the middle category and significant. Yet, for bottom positions Income is now insignificant and the relative risk of being in the bottom positions are 1,249 times more likely than obtaining a middle position.

Given a unit increase (tenth of thousand £) in Population, the relative risk of being in the top positions are 1,005 times more likely than being in the middle positions. Furthermore, the relative risk of being in the bottom positions, given a ten million £ increase in Population, are 0.997 times more likely than being in the middle positions, but not significant. However, note that the p-values for both top and bottom positions have decreased drastically in comparison with multinomial logit regression #1.
7. Discussion

Regarding Turnover, the results have displayed positive effects on final position. The predictor is intact throughout the regressions and has proven as a significant measure of final position. Furthermore, these results are in line with the presented hypothesis. It could be argued that the marginal utility for a ten million £ increase in Turnover (given that all of the other variables are held constant) for both top and bottom teams are equal since their respective coefficients provide a similar effect on final position (see table 5).

Broadcasting revenues is part of turnover and top tier clubs in the EPL receive a larger share of the broadcasting revenues at the expense of clubs with less financial capabilities. Thus, this stream of income is distributed unevenly. Ultimately, this situation endangers the uncertainty of outcome principal and therefore the attractiveness of the EPL, since fewer teams have a plausible possibility to win the league title (Downward & Dawson, 2000). Arguably, this would increase the Gini-coefficient, that represents a more unbalanced league, in accordance with Curley & Roeders’ (2016) argument. Tentatively, if the distribution of broadcasting revenues would be more equal this would increase the turnover for bottom teams, increase the uncertainty of outcome and increase the attractiveness of the EPL.

As previously mentioned in the ‘Limitation’ section, home league match attendance is also part of Turnover. Since the underlying reason is unknown for home league match attendance whether attendance is the reason behind success or a result of success, suspicions regarding endogeneity problems arises. In this instance, whether Position effects Turnover or Turnover effects Position. However, it could be argued that in the EPL, home league match attendance is not dependent on success since almost every fixture is sold out (the price elasticity for attending a game is very low). Thus, endogeneity concerning Turnover may be rejected.

Initially, the variable Previous was utilized when analyzing previous year’s final position. Previous displayed a negative and significant effect on Position which opposes the presented hypothesis. However, by dividing Previous into separate effects (dummy variables) a more accurate measure was created. Now, each tier in the league receives a more justified result since middle teams seem to fluctuate more than top and bottom teams when observing previous league tables. Ultimately, the dummy who opposed Previous’s predicted effect was
DummyTop, DummyBottom and DummyPromotion were consistent with the result displayed by Previous.

The coefficients for DummyTop, DummyBottom and DummyPromotion displayed consistent results throughout all the regressions. Furthermore, the results are in line with our hypothesis that top teams the previous year will remain in the top, and bottom and promoted teams will stay in the bottom or be relegated. The results for these variables, except DummyTop for top positions, are also significant throughout the analysis which states that the effect is trustworthy. It would appear that this conclusion is logical since there is a clear top six in the EPL every year and that promoted teams struggle to establish themselves. Notably, promoted teams have a higher probability of remaining in the EPL than bottom teams. Arguably due to momentum from a long winning streak the past season. On the contrary, bottom teams originate from a season with poor results, which may affect their momentum negatively and ultimately their final position. Conclusively, top teams the previous year have never obtained bottom positions the following year and vice versa for bottom and promoted teams. This should explain the high significance level for these predictors.

Competition received a negative effect on final position which opposes the stated hypothesis. Furthermore, competition presents high p-values throughout the regressions which indicates difficulties to draw any valid conclusions. It is therefore questionable to include in the final model. A possible reason behind this could be that the only clubs with substantial rivalry are the London clubs. Moreover, these clubs obtained diverse final positions throughout all of the observed seasons which may be the reason that this variable is complex to evaluate.

Regarding MeanTemperature, both non -and significant values have been observed. Additionally, this variable displays high correlation with other independent variables (see table 2). The hypothesis that warmer temperatures provide better prerequisites and thus a positive effect on final position is consequent with the results. However, the influence it provides is greater than expected. Beliefs that one-degree Celsius increase in mean temperature should affect final position to such an extent may not be believable since the mean temperature in the UK does not vary that much between regions. It will therefore be excluded from the final model.

The hypothesis for the predictor Income was that it would affect final position positively. However, the result for the parameter Income had a negative impact on the dependent variable.
in both the ordered and multinomial logit regressions. This implies that teams experience a negative effect of originating from regions with high Income, e.g., the clubs from London who are inconsistent in their performances. Another possible reason behind this negative effect is that clubs that originate from regions with low Income have obtained top positions e.g., Manchester City, Manchester United and Liverpool. Furthermore, it could be argued that individuals from high Income areas have different preferences and that they may substitute football in exchange for other commodities. In conclusion, Income has a negative impact on clubs’ utility maximization function.

Population has a positive effect on position in all regressions and is in accordance with the stated hypothesis. However, the only time it is significant is for top positions in multinomial logit regression #2. Note that Population provides low coefficients which indicate a marginal effect on final position. As observed in the correlation matrix, Competition and Population are correlated, since cities with large population have a high degree of competitiveness due to many clubs and vice versa. Yet, even though Population provides a marginal effect, it might validate the model since it makes other variables more significant and may also act as a measure of Competition. Arguably, if Population in an area increases, demand for the local club will too, which is also demonstrated by El-Hodiri & Quirk (1971) and Szymanski & Kuypers (1999). This generates a low-price elasticity on match day tickets and thus more revenue. Additionally, with increased demand, a consequence can be both positive and negative external effects in terms of social links and territorial dynamics. In the ‘External effects’ section these positive effects can appear in fashions like increased fan base and stronger incentives to invest in the club. Furthermore, the negative ones could appear in terms of hooliganism and crowding-out effects.

The predictor NetTransfers displayed significant negative effects on final position concerning the two ordered logits and the top positions for the multinomial logits. This indicates that if clubs purchase more than they dispose, in terms of players, it will affect final position negatively. This result opposes the stated hypothesis. The question seems to be whether additions make an instant impact or not. Therefore, an attempt of lagging NetTransfers with one year was performed where it affected Position positively, however without any mentionable degree of significance.
Possible reasons to why purchasing more would suffer the club’s final position could be hasty or misinformed investments, either through poor scouting or ignorant decision makers. According to Soderman (2012), there is an interface between managers and the board of directors on what acquisitions that needs to be made. However, the board has authority in the final decision. Arguably, this interplay between manager and board could be inadequate to the extent that the board takes all transfers decisions on their own. Furthermore, the manager tends to possess more knowledge regarding these matters since they observe the players on a daily basis. Perhaps, clubs with poor interplay between manager and board is ordinarily and may thus affect final position negatively.

Soderman (2012) continue his reasoning by stating that buying players is of vital importance. Therefore, the decision needs to be of a strategic fashion, that is, thinking long-term. However, thinking long-term is problematic since individuals in senior positions within a football club runs the risk of having their contracts terminated if results are poor. Thus, planning for the long-term is difficult if individuals in key position do not have security of tenure. Ultimately, the board tends to purchase expensive, experienced and more tested players to achieve immediate success (Moore & Levermore, 2012). It appears that a perfect acquisition strategy is difficult to achieve since balancing short-and long-term thinking is a complex matter. This complexity may also be a reason behind the negative effect NetTransfers has on Position.

SquadValue displayed a positive impact on final position with significant values. Although, the effect on final position was minor. As expected, these results were in line with the stated hypothesis. Described in the data section, SquadValue represents the capacity and skill of all the players within a club. Moreover, a higher squad value may indicate a broader squad which provides opportunities for the manager to elaborate and optimize the team. It could also be argued that it is important when there are plenty of games played within a short period, where clubs in the EPL participate in several different championships, ranging from domestic to international cups. Higher SquadValue may therefore result in more resistance to injuries and more players in good condition. Another possible aspect to why SquadValue affects Position positively could be that it creates a degree of competition that benefits the club’s performance. Where players need to perform to their utmost capability in order to play regularly. However, the predictor SquadValue will be excluded from the final model due to high correlation with Turnover and low influence on final position.
8. Conclusion

Initially, attempts on an ordered logit regression was performed since the independent variable Position demonstrates a clear order ranging from Bottom to Middle to Top. However, the Proportional Odds Assumption was eventually violated which brought the analysis to a multinomial logit regression with Middle as the reference group. Separate effects could therefore be observed on Top and Bottom for each predictor. Furthermore, a pervading emphasis in this thesis has been the assumption that final position is a vital component for the perceived utility a club experience. Therefore, potential factors that could affect Position have been explored and analyzed. Ultimately, the following variables are deemed as interesting and thus relevant to include in the final utility maximization function; Turnover, DummyTop, DummyBottom, DummyPromotion, Income, Population and NetTransfers. The final adjusted model, in accordance with Sloan’s (1971) model, is presented as:

\[ U_{it} = \alpha_{it} + P_{it}(\beta_{Turnover_{it}} + \beta_{Income_{it}} + \beta_{Population_{it}} + \beta_{NetTransfers_{it}} + \beta_{DummyTop_{D1_{it}}} + \beta_{DummyBottom_{D2_{it}}} + \beta_{DummyPromotion_{D3_{it}}} )X_i(\pi_{ri_{it}} - \pi_{0_{it}} - T_{it}) + \epsilon_{it} \]

\[ D1_{it} = \begin{cases} 1 & \text{Top} \\ 0 & \text{Others} \end{cases} \]

\[ D2_{it} = \begin{cases} 1 & \text{Bottom} \\ 0 & \text{Others} \end{cases} \]

\[ D3_{it} = \begin{cases} 1 & \text{Promotion} \\ 0 & \text{Others} \end{cases} \]

The utility perceived for each club (U) depends on position (P) which itself depends on Turnover, DummyTop, DummyBottom, DummyPromotion, Income, Population and NetTransfers. The utility also depends on the degree of competitiveness (X). Furthermore, the utility perceived by each club is subject to the financial constraint \((\pi_{ri_{it}} - \pi_{0_{it}} - T_{it})\). Finally, the indexation \(i\) indicates that the variable is specific for 34 teams and \(t\) that the variable is specific over 9 years.

Income and NetTransfers are the only variables in the final model that differ from previous research in terms of the effect on Position (Carmichael, McHale & Thomas, 2010; Leeds & Leeds, 2006) As stated before, this thesis has observed a negative effect from both Income and NetTransfers.
Moreover, the parameters in the final model that affected Position the most was DummyTop for top positions and DummyBottom for bottom positions. This highlights the inequalities in the EPL that has emerged in the recent decade. Teams are now playing sub championships, where there is a defined top six who are competing over the Champions League spots. This new paradigm of the EPL endangers the uncertainty of outcome principle and therefore the attractiveness of the EPL, where there are a lot fewer teams which have a plausible chance of winning.

Another conclusion worth mentioning is that the EPL, as a commodity, could be argued as under-produced. Reasons being a low-price elasticity and neglection of social costs or benefits that occurs when an external effect becomes present, e.g., increased fan base or revenues generated by the media, that derives from club activities. Additionally, this might explain why Population affects Position positively in the sense that it derives from an increased fan base that supports the team.

The conclusion from this analysis is that clubs maximize their utility (obtaining a higher final position) by increasing turnover and building on previous success. Additionally, top teams ought to be careful when purchasing players since it seems to affect their final position negatively. Simultaneously, bottom teams may experience greater incentives to purchase players to achieve better results the upcoming season. Since Income and Population are external factors that the clubs cannot affect directly, they become complicated to manage. Yet, these variables should be considered as they affect final position according to the final model.
9. References


Soderman, S. (2012) Football and Management: Comparisons Between Sport and Enterprise retrieved from: https://books.google.se/books?id=Rb0TIPYzYZ4C&printsec=frontcover&hl=sv&source=gbs_ge_summary_r&cad=0#v=onepage&q&f=false


Appendix A

For the 2008/2009 season Portsmouth finished 14th in the EPL, which assign them into the second category of the dependent variable (position 5-17). For this season’s calculation of Turnover, Net Transfers and Squad Value every team’s individual value, for each variable, has been summarized and then divided by number of teams. This is then repeated for the 2009/2010 season but with the third category of the dependent variable (position 18-20), as Portsmouth’s final position was 20th in the EPL. This provides Portsmouth an average amount for each season to recreate a somewhat valid number.

Appendix B

For the calculation of Income for 2017 an assumption of constant economical change is made. The estimation for regional gross disposable household income per head for 2017 is calculated as; Every region’s annual percentage increase or decrease in income, summarized, and then divided on the number of years observed for each region, to get the average percentage change for each region. The average percentage change for each region times gross disposable household income per head for 2016 provides the value for the explanatory variable Income for 2017 for each region.
Appendix C

Table 6, Ordered logit regression #2

<table>
<thead>
<tr>
<th>Independent Variable</th>
<th>$\beta$</th>
<th>Robust S.E</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>DummyTop</td>
<td>2,950</td>
<td>2,346</td>
<td>0,174</td>
</tr>
<tr>
<td>DummyBottom</td>
<td>0,188***</td>
<td>0,096</td>
<td>0,001</td>
</tr>
<tr>
<td>DummyPromotion</td>
<td>0,314**</td>
<td>0,165</td>
<td>0,028</td>
</tr>
</tbody>
</table>

Proportional Odds Assumption | $p>\text{chi2}$ | 0,000 |

Likelihood Ratio Test        |          |       |

* Significance level 10 %    ** Significance level 5 %    *** Significance level 1 %

Notable mentions for independent variables, given that all of the other variables are held constant, in ordered logit regression #2:

For all independent variables except the Dummy ones, minor changes occurred compared to ordered logit regression #1 which do not change the interpretations of the odds ratios in any substantial way. For top teams the previous year, the odds of a top position versus the combined middle and bottom positions are 2,950 greater than for middle, bottom and promotion teams. Note that DummyTop is only significant on a 20% level. For bottom teams the previous year, the odds of observing a top position versus the combined middle and bottom positions are 0,188 times lower than for top, middle and promotion teams. For promotion teams the previous year, the odds of observing a top position versus the combined middle and bottom positions are 0,314 lower than for top, middle and bottom teams.