ERASMUS UNIVERSITY ROTTERDAM

Erasmus School of Economics

Master Thesis

Reference-Dependence in Dutch Football

Outcome Uncertainty and Stadium Attendance

Author: Martijn Oeij
Student number: 346271
12-06-2017

Supervisor: Prof. dr. Han Bleichrodt
Co-evaluator: Timo Lambregts

Abstract

The purpose of this thesis is to look into the relationship between uncertainty of the outcome of Dutch football games and the stadium attendance of those particular games. Two main theories are proposed by previous literature as to what this relationship looks like. The 'classic' uncertainty of outcome hypothesis (UOH) suggests that more uncertainty yields more stadium attendance. A model with reference-dependent preferences coupled with loss aversion proposes the opposite, i.e. less uncertainty yields more stadium attendance. Three Dutch football clubs provided attendance data for the analysis of this thesis, namely FC Groningen, Heracles and PSV. Betting odds are used to calculate the uncertainty of outcome. The estimated random effects tobit model provides evidence in favour of the UOH. The maximum stadium attendance is achieved at the uncertainty level where the home team is three times as likely to win as the away team. Other significant influences on stadium attendance are the away team, the league rank of the away team, derby matches, renovation of the stadium, the day of the game, the month in which the game is played, and the season.

Acknowledgement

I would first like to thank my master thesis supervisor prof. dr. Han Bleichrodt of the Erasmus School of Economics. Prof. Bleichrodt helped me throughout the challenging process of writing this thesis. He steered me in the right direction whenever I had trouble with anything. Data collection was perhaps the most challenging part and prof. Bleichrodt supported me with approaching football clubs. Despite the fact that a relatively small number of football clubs were willing to provide data, prof. Bleichrodt acknowledged my enthusiasm for the subject and allowed me to pursue my own interest with great freedom in my approach.

I would also like to thank the football clubs that were willing to participate in the research of this thesis. In particular thanks to Coen Schroots from PSV, Martijn van der Lee from Heracles, and Richard van Elsacker from FC Groningen. In a busy Eredivisie season they found the time and willingness to provide me with data.

Table of contents

1. Introduction

Uncertainty can be an important component of what makes a sporting event interesting. A team or athlete never has a 100 per cent probability of winning and this is part of the fun. In sports economics, extensive research has been done that looks into the relationship between uncertainty in the outcome of a sports game and the demand for attending this particular game. The uncertainty of outcome hypothesis (UOH) states that in a particular game uncertainty about the outcome influences the demand for the sports event in question. More specifically, the UOH argues that more uncertain games attract more consumers. The literature investigating the UOH dates back to a 1956 paper by Rottenberg (1956). He argues that an equal distribution of playing talent across Major League Baseball teams grants uncertainty of outcome, and uncertainty of outcome or competitive balance is the reason consumers pay admission fees to see these teams play. Neale (1964) also acknowledges that in sports the excitement comes from changes in the league standings of sports teams, and that more frequently changing league standings would lead to more gate receipts. He adds an insightful paradox in favour of the UOH, which is called the Louis-Schmelling Paradox, named after two former boxers. This paradox describes the situation in which one team or athlete has a monopoly in the market. A traditional firm would like to have a monopoly for its maximising profits. However, in sports this is different. If, for instance, the New York Yankees would contract all the good players from other American baseball teams, this would mean that these other teams would fail in establishing a (competitive) team. In Neale's (1964) words this would mean "no games, no gate receipts, no Yankees". The Louis-Schmelling paradox is therefore a striking example to show that even when a sports team wishes to become the absolute best, they still need competition in order to generate revenue.

It took some years before researchers actually tried to find empirical evidence for the UOH. An extensive literature review by Coates et al. (2014) provides a clear overview of the various articles that discuss the UOH in more debt. The articles in question differ in what sport they investigate, as well as their modelling approach. Australian football, football (soccer), American football, baseball, rugby and hockey are common sports where research is done exploring the UOH. The main conclusion from Coates et al. (2014) is that the UOH is controversial. The authors develop a model with reference-dependent preferences and with their own data from the Major League Baseball, the authors show that the UOH does not hold. In fact, games with less uncertainty yield higher attendance figures.

The model with reference-dependent preferences developed by Coates et al. (2014) is relatively new and can be applied to other sports. In short, reference-dependent preferences combined with loss aversion predict an opposite relation between uncertainty of the outcome and attendance of a sports match, compared to the UOH. Therefore, it is interesting to see how this reference-dependent preference model can help in identifying the UOH in football (soccer). Research is done on the topic of UOH and European football leagues. However, none have looked into Dutch football. This thesis will focus on the highest Dutch national football league, called the Eredivisie. The subject of competitive balance is still being looked at in the Dutch leagues. In May 2017 Dutch football clubs rejected several proposals for reforming the league (NU, 2017), including a proposal to reduce the number of participating clubs in the highest division. One of the arguments for reforming the Eredivisie is competitive balance and the fact that this would yield a higher attendance rate. However, results from previous studies regarding the UOH are mixed and therefore research needs to be done to see if competitive balance is indeed an argument for reforming the league to attract more followers. This leads to the following research question:

Does uncertainty in the outcome of a match increase or decrease the stadium attendance for clubs in the Dutch Eredivisie?

To get insight into this topic, the UOH model with reference-dependent preferences is outlined and explained in detail. Secondly, the relevant literature is discussed. Finally, the model is applied to data from several clubs playing in the Eredivisie. With betting odds and attendance data from several Dutch football clubs the relationship between the probability of winning and live game attendance is analysed. The following sections elaborate on these research steps.

2. The model

Coates et al. (2014) developed a model for investigating in which way uncertainty in the outcome of a particular sports game influences stadium attendance. They extend the model by Card and Dahl (2011), who look at unexpected outcomes in American football and the relationship with family violence. Coates et al. (2014) further optimize this model to investigate the UOH. This model looks exclusively at individual match level uncertainty of outcome and this will also be the focus in this thesis. The model is based on the assumption that a game has two outcomes, namely a win or a loss. In the remainder of the explanation of the model this assumption is extended to football, while in reality a draw is a common outcome. In the dataset used in this thesis 20.53% of the games ended in a draw. This issue is discussed later in this section.

According to standard consumer theory, consumers who attend a game receive "consumption utility". The utility gained from a win is given by U^W and the utility gained from a loss is U^L, where

$U^W > U^L$ is assumed. The objective probability of winning a game for the home team is given by p, where $0 \leq p \leq 1$. Expected utility from attending a game for home supporters is then given by

$$E[U] = pU^W + (1-p)U^L. \quad (1)$$

In this standard consumer theory model expected utility increases with the probability of a home win. Following this model it is expected that more successful teams have higher attendance. However, this is not consistent with the UOH. Furthermore, from prospect theory it is known that reference points play an important role in decisions under risk (Kahneman & Tversky, 1979). Therefore, the standard consumption utility model needs an extension. Koszegi and Rabin (2006) call this extension "gain-loss utility", which is the part of the utility function that examines the relationship between the actual outcome of an event and the reference point for that particular event. Each event has its own reference point.

Assume that the outcome of a football match y equals 1 if the match is won by the home team and 0 if the home team loses. The reference point for the home fan for the football game is given by $E(y=1) = p^r$. Deviation from this reference point generates gain-loss utility, according to the model developed by Koszegi and Rabin (2006). The marginal impact of this deviation is α if the home team wins, where $\alpha > 0$. The utility for home fans if their team wins is given by

$$U^W + a(y - p^r) = U^W + \alpha(1 - p^r). \quad (2)$$

From this equation it can be seen that a positive deviation from the reference point results in an increase in total utility, i.e. an unexpected win results in higher utility compared to an expected win. The marginal impact in the gain-loss utility function when the home team loses is given by β, where $\beta > 0$. The utility for home fans if their team loses is given by

$$U^L + \beta(y - p^r) = U^L + \beta(0 - p^r). \quad (3)$$

This equation shows that a negative deviation from the reference point results in a decrease of total utility, i.e. an expected loss generates higher utility compared to an unexpected loss.

Figure 1 provides a graphical illustration on the relationship between the match outcome, reference points and the utility generated by those two factors. It can be seen that the highest utility is achieved when $p = 0$, i.e. the home team is perceived to have 0 per cent probability of winning, and the home team wins. Likewise, the lowest utility is generated when $p = 1$, i.e. the home team is perceived to have 100 per cent probability of winning, and the home team loses. These two extreme cases are the scenarios in which the deviation from the reference point in the gain-loss utility is at its maximum.

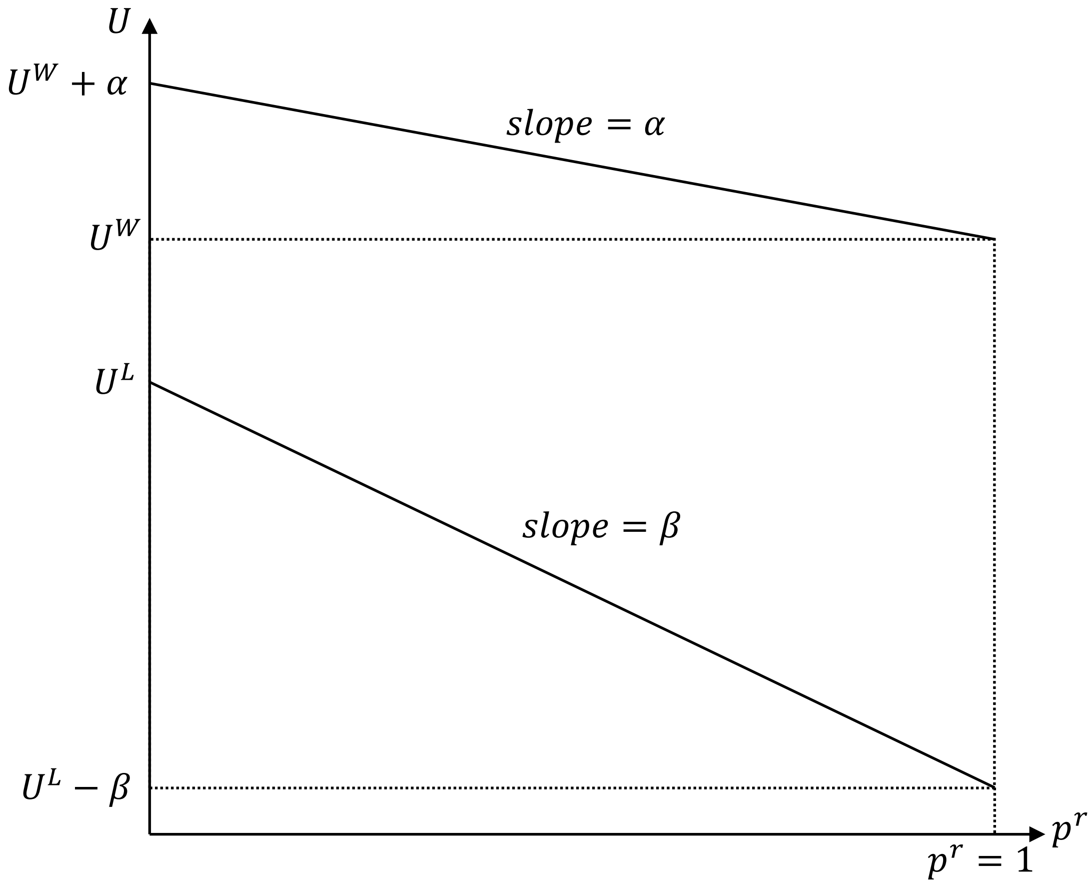

Figure 1 Utility of game outcomes with references

Now assume the reference point for the home fans equals the objective probability that their team wins. Utility from attending a game is given by

$$E[U] = p[U^W + \alpha(1-p)] + (1-p)[U^L + \beta(0-p)]. \tag{4}$$

After rearranging the following equation is obtained

$$E[U] = (\beta - \alpha)p^2 + [(U^W - U^L) - (\beta - \alpha)]p + U^L. \tag{5}$$

This utility equation incorporates both consumption utility (equation 1) and gain-loss utility and shows a quadratic function of the probability that the home team wins.

Consumers use the utility function from equation 5 in deciding whether or not to attend a game. They compare the expected utility from equation 5 to a reservation utility v that they get when not attending a game, i.e. the reservation utility represents other leisure activities. This reservation utility v has a distribution of $[\underline{v}, \overline{v}]$. The most determined fans will have low reservation utility close to $\underline{v}$, while casual fans have a higher reservation utility. This means that for casual fans a higher expected utility is needed before they decide to attend a game.

2.1 The Uncertainty of Outcome Hypothesis

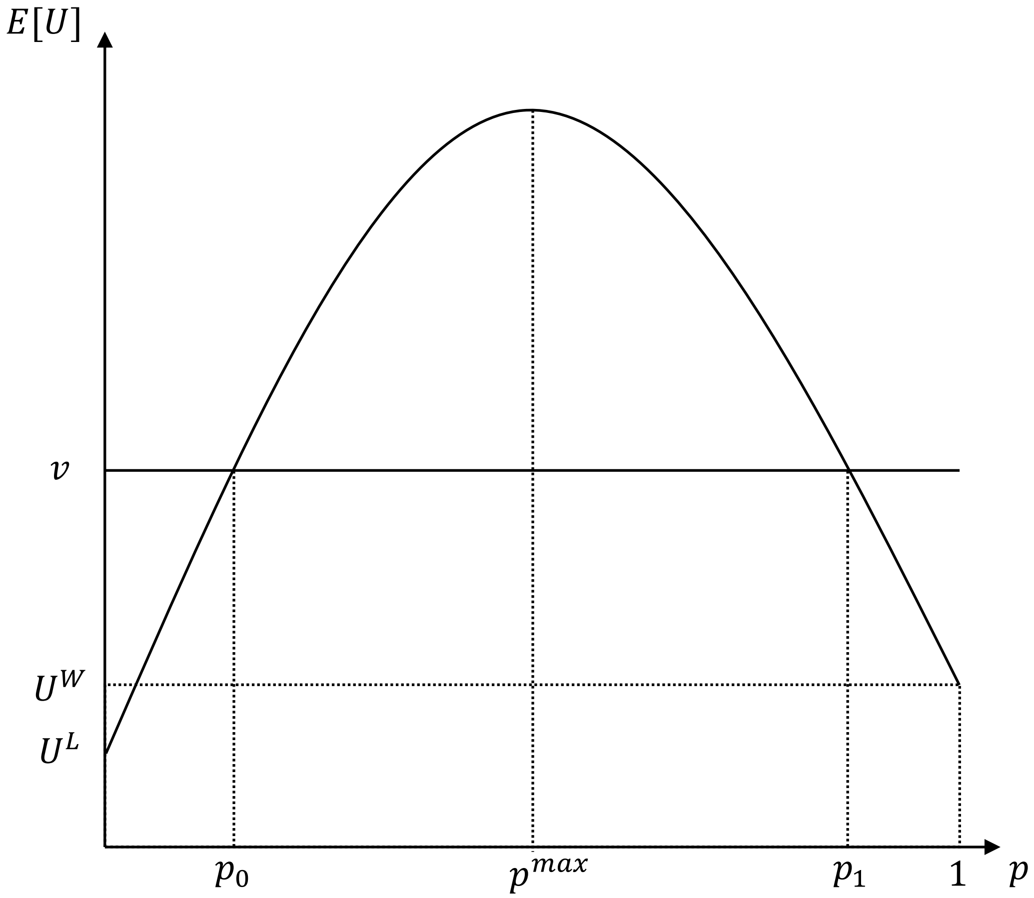

Figure 2 Classic UOH, concave function of probability

As named by Coates et al. (2014), the "classic" UOH predicts a concave relationship between the probability of the home team winning and the expected utility of attending a game. This is illustrated in figure 2. The maximum p^{max} is reached at the probability interval (0.5, 1) of the home team winning, likely around 55 per cent as hypothesized by Rottenberg (1956). In figure 2 it is easy to see that as the outcome becomes more uncertain, i.e. the probability of winning gets closer to p^{max}, more consumers along the distribution of reservation utility $[\underline{v}, \overline{v}]$ will decide to attend a game. In the case presented in figure 2, the consumer will decide to attend the game if $p_0 < p < p_1$. Concavity of the expected utility function is achieved if in equation 5 $(\beta - \alpha) < 0$ and $[(U^W - U^L) - (\beta - \alpha)] > 0$. The first condition $(\beta - \alpha) < 0$ that needs to hold for the classic UOH states that the marginal impact of a positive deviation from the reference point is greater than the marginal impact of a negative deviation from the reference point. In other words, an unexpected win should result in higher marginal utility than an unexpected loss. The second condition $[(U^W - U^L) - (\beta - \alpha)] > 0$ is logical, given that $(\beta - \alpha) < 0$ and $U^W > U^L$.

Using the first derivative of equation 5 it is possible to define p_{max}. The first order condition is defined as follows

$$p^{max} = \frac{1}{2} - \frac{(U^W - U^L)}{2(\beta - \alpha)} \tag{6}$$

where, given that $(\beta - \alpha) < 0$, it can be shown that $p^{max} > \frac{1}{2}$. Also the classic UOH is subject to $p^{max} < 1$ by definition, which combining with equation 6, shows that $(U^W - U^L) < (\alpha - \beta)$. This relationship shows that a consumer prefers more uncertain games over their team winning for sure. This is exactly what the classic UOH hypothesizes and in this reference-dependent preference model this is the way in which the classic UOH is justified.

2.2 Loss aversion

Contradictory to the UOH is a situation in which more certain games are preferred over uncertain games. This case can best be described by loss aversion. As Kahneman and Tversky put it: "losses loom larger than gains" (Kahneman & Tversky, 1979). In other words, losing results in a bigger decrease in utility than a gain leads to an increase in utility. In the sports framework presented in this thesis this would translate into $\beta > \alpha$, meaning that the marginal impact of a negative deviation from the reference points is larger than the marginal impact of a positive deviation from the reference point. Instead of the concavity of the classic UOH utility function, the utility function under loss aversion is convex, as is shown in figure 3.

In order to separate the consumption utility and the gain-loss utility in the case of loss aversion, equation 4 is rearranged into

$$E[U] = [pU^W + (1-p)U^L] + (\alpha - \beta)p(1-p). \tag{7}$$

The consumption utility $[pU^W + (1-p)U^L]$ increases with p. The gain-loss utility $(\alpha - \beta)p(1-p)$ first decreases with p until $p = \frac{1}{2}$, then increases with p, given that $(\alpha - \beta) < 0$ for loss aversion to hold. The lowest utility is achieved at p^{min} and the first order condition is now $p^{min} = \frac{1}{2} - \frac{(U^W - U^L)}{2(\beta - \alpha)}$, which is similar to equation 6. If $p < p^{min}$ then the negative effect of the gain-loss utility outweighs the positive effect by the consumption utility. When $p > p^{min}$ the opposite is the case. In short, with reference-dependent preferences and loss aversion, attending a game that is either an expected win or an expected loss, generates more utility than games with more uncertainty. So, in this model an expected loss, e.g. $p < 0.1$, is preferred to a more uncertain match, e.g. $p = 0.5$. This can be motivated by the fact that casual fans are interested in seeing surprising results, i.e. 'David versus Goliath' scenarios (Buraimo, 2014). This happens when the home team is highly expected to lose, e.g. $p < 0.1$, and the home team wins regardless of that probability. These surprising results are not

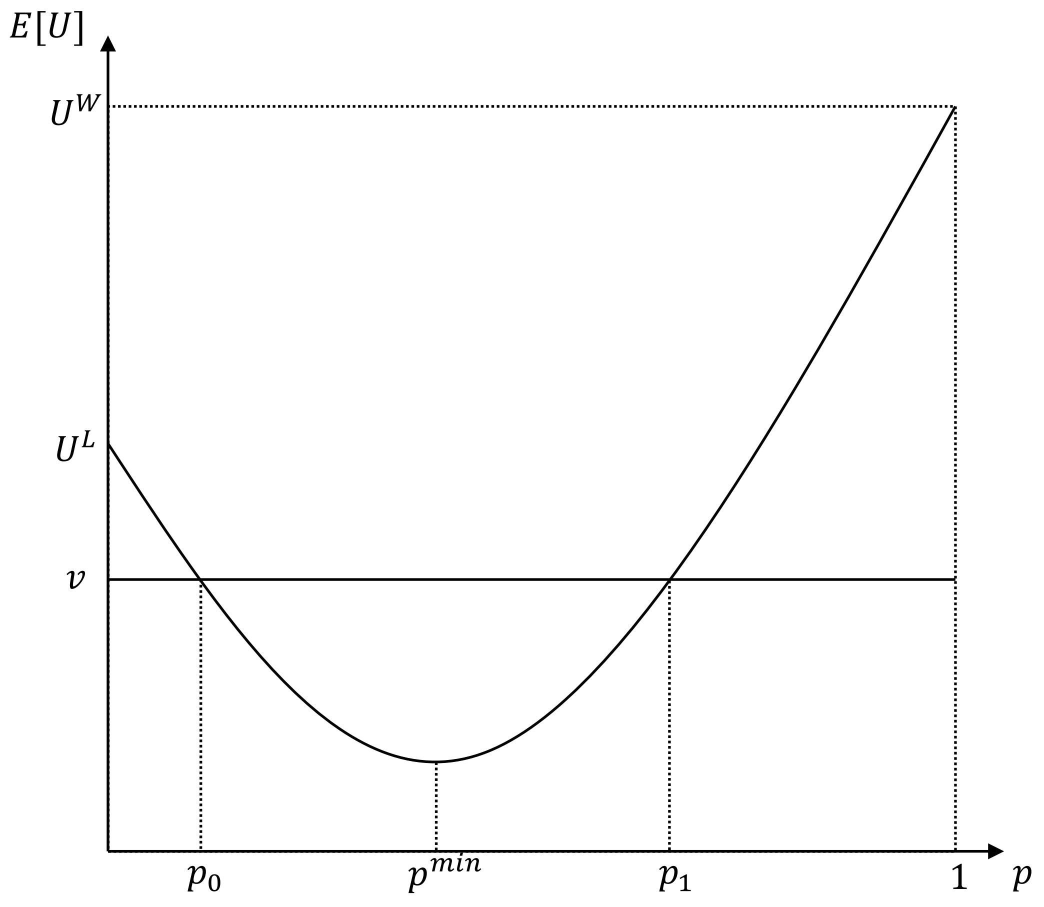

Figure 3 Reference-dependent preferences and loss aversion

explained in the classic UOH framework, while in reality it could be the case that some fans seek games to attend at which they can witness a potential surprising result. Also, this version of the model allows for explaining the fact that fans would only want to see their team win, something the classic UOH fails to recognise.

In figure 3 p_0 and p_1 are the cut-off points for which attending the live game is more attractive than not, given reservation utility v. In the case of loss aversion the reservation utility is important. If for some consumers the reservation utility is relatively high, then $v > U^L$. In this situation p_1 is at a relatively high level and therefore identification of the utility function with loss aversion is empirically difficult. The reason for this is that in the situation where $v > U^L$, observations where $p < p_1$ may show a flat relationship between utility and p. If the reservation utility is relatively low, which is assumed to be the case for strong fans of a team, then this will not be a problem. Furthermore, the difference between U^W and U^L is crucial. If this difference is relatively large and with $U^W > U^L$, then the expected utility function will have the same relationship to p as in equation 1, i.e. increasing in p.

2.3 Application to football

This framework assumes that the outcome of a game is binary, i.e. $y = 1$ for a win and $y = 0$ for a loss. However, in football the outcome of a draw is not uncommon. In league football a win results in 3 points for the winning team and a draw results in 1 point. This complicates the model. The solution for this is as follows. Suppose fans of a strong team expect their team to win against a weak opponent. In this case even a draw for the fans of the strong team feels like a loss; they 'lose' 2 points when their reference is winning. So for strong teams the solution in the model is to transform the binary variable into two outcomes: a win $y = 1$ and not winning $y = 0$ (draw or loss). Hereby, it is the assumption that in this case the utility of a draw and loss is the same, which is a questionable assumption. For a weak team the reasoning is the same. Fans of a weak team would like a draw against stronger opponents. Therefore, in the outcome of a draw they 'win' 1 point. For weak teams the binary variable y becomes 1 when winning or drawing, and 0 when losing.

These assumptions are open for debate, because it is questionable to assume for fans of a strong team the utility generated by a draw equals the utility generated by a loss. However, for empirically looking at data and interpreting the results, this does not matter. For the reference-dependence models with either classic UOH or loss aversion to hold, the relationship between the probability of winning and stadium attendance remains the same, despite the fact that a third outcome could be added to the model. Also, for the analysis of data in the following sections, the fact that the draw outcome complicates the theoretical model does not affect the implications or interpretations of the results. In the literature section different studies investigating football are mentioned which all incorporate a form of p. The authors from these studies all compare their results to the same classic UOH as authors from other sports.

2.4 Econometric model

Looking at the expected utility equation 5 it is possible to simplify the terms as a function of the probability of the home team winning. This looks like

$$E[U] = \gamma p^2 + \theta p + \lambda. \tag{8}$$

Now it is possible to derive an econometric model (Coates et al., 2014) explaining attendance at live games. Therefore, it is assumed that in the observations the attendance of a game depends on the number of people that have higher expected utility of attending a game compared to their reservation utility, so $E[U] > v$. The econometric model with home team i and away team j at time t is as follows

$$\ln Attendance_{ijt} = \lambda + \gamma p_{ijt}^2 + \theta p_{ijt} + f(covariates) + \varepsilon_{ijt} \tag{9}$$

where $f(covariates)$ are parameters for characteristics of the home team, the away team, the match itself, and time variables, and ε_{ijt} is the random error. The parameters of interest are γ, θ and λ, where $\gamma = \frac{\beta-\alpha}{\overline{v}-\underline{v}}$, $\theta = \frac{(U^W-U^L)-(\beta-\alpha)}{\overline{v}-\underline{v}}$ and $\lambda = \frac{U^L-\overline{v}}{\overline{v}-\underline{v}}$ according to the framework developed in the previous sections. With this econometric model it is now possible to look at other literature for results concerning the influence of the probability of winning for the home team on live attendance. In order to do this in a well-ordered way, it is useful to distinguish between the different scenarios and match these scenarios with the reference-dependent preference models with classic UOH or loss aversion. The following hypotheses are therefore needed (Coates et al., 2014):

- *H1* Loss aversion is supported by the model: $\gamma > 0$, i.e. $\beta > \alpha$.
  - *H1a* $\gamma > 0$ and $\theta < 0$, i.e. $\beta > \alpha$ and $(U^W - U^L) < (\beta - \alpha)$. This means that for the marginal consumer the marginal impact of loss aversion is larger than the consumption utility difference between a home win and a home loss, given that $U^W > U^L$.
  - *H1b* $\gamma > 0$ and $\theta > 0$, i.e. $\beta > \alpha$ and $(U^W - U^L) > (\beta - \alpha)$. This means that for the marginal consumer the marginal impact of loss aversion is smaller than the consumption utility difference between a home win and a home loss, given that $U^W > U^L$.
- *H2* The marginal consumer does not behave according to reference-dependent preferences and the utility for a home win is larger than the utility for a home loss: $\gamma = 0$ and $\theta > 0$, i.e. $\beta = \alpha$ and $U^W > U^L$.
- *H3* The classic UOH is supported by the model: $\gamma < 0$ and $\theta > 0$, i.e. $(\beta - \alpha) < 0 \leq (U^W - U^L)$. This means that the marginal consumer has a preference for uncertain games and gets more utility from an unexpected win compared to an unexpected loss.

With these structured hypotheses it is possible to look at the existing literature, and conduct an analysis with data from Dutch football clubs.

3. Literature

An overview of the relevant literature that investigates the UOH in football leagues is presented in table 1. These papers all include a measure for outcome uncertainty as an independent variable and stadium attendance as the dependent variable. Furthermore, most papers use a similar approach as to what covariates to use in the estimation of the econometric model. These covariates often include variables for quality and form of both teams, and certain incentive variables, such as the stage of the competition, whether a game is considered a derby, distance between the two teams, and time, day and month of the game. In the remainder of this section these papers are discussed more elaborately.

The first notable papers are by Peel and Thomas (1988; 1992). They investigate the UOH in the top four divisions of the English Football League[1]. In their 1988 paper they include only a linear measure for outcome uncertainty, namely the probability of winning for the home team, derived from betting odds. Also, they employ data on the position of the home/away team, the distance between the home and away team, a variable for derby matches, and a variable indicating what stage the competition is at. With these parameters they estimate an ordinary least squares (OLS) model, which is a relatively simple way of looking at the UOH, compared to the literature that is later discussed. Peel and Thomas (1988) find with this simple model that their single variable for home team probability of winning is positive and significant for all four divisions separately, meaning that

Author(s)	**Football league**	**Uncertainty measure**	**Results**	**Support**
Peel and Thomas (1988)	English tier 1-4 1981-1982	Betting odds	$\theta > 0$	H2
Peel and Thomas (1992)	English tier 1-4 1986-1987	Betting odds	$\theta < 0$, $\gamma > 0$	H1a
Czarnitzki and Stadtmann (2002)	German top tier 1996-1997	Betting odds	$\theta = 0$, $\gamma = 0$	-
Forrest and Simmons (2002)	English tier 2-4 1997-1998	Betting odds	-	H3
Forrest et al. (2005)	English tier 2-4 1997-1998	Betting odds	-	H3
Falter, Pérignon, & Vercruysse, 2008	French top tier 1996-2000	$f(points)$	$\theta = 0$	-
Buraimo & Simmons (2008)	English top tier division 2000-2006	Betting odds	$\theta < 0$, $\gamma > 0$	H1a
Benz et al. (2009)	German top tier 2001-2004	Betting odds, $f(win\%)$	$\theta > 0$, $\gamma < 0$	H3
Madalozzo & Berber Villar, 2009	Brazilian top tier 2003-2006	$f(win\%)$	$\theta = 0$	-
Buraimo and Simmons (2009)	Spanish top tier division 2003-2007	Betting odds	$\theta < 0$, $\gamma > 0$	H1a
Buraimo (2014)	English tier 1-5 2006-2012	Betting odds	$\theta < 0$, $\gamma > 0$	H1a
Reilly (2015)	Irish top tier 2012-2014	Betting odds	-	-
Martins and Cró (2016)	Portuguese top tier 2010-2015	Betting odds	$\theta < 0$, $\gamma > 0$	H1a
Jena and Reilly (2016)	Irish 2nd tier 2013-2015	Betting odds	$\theta > 0$, $\gamma < 0$	H3

Table 1 Literature of outcome uncertainty and live game attendance in football

[1] The well-known English Premier League was formed in 1992, so before 1992 the top tier of the English Football League was the highest division in English football.

attendance is higher when the home team has a greater chance of winning (H2). Their subsequent paper further develops this simple model, with two extensions, namely a variable for the square of the probability of the home team winning, and a measure of the attendance of the preceding match for the home and away team. Using, once again, an OLS model, the results by Peel and Thomas (1992) suggest a relationship between live game attendance and outcome uncertainty that supports H1a.

Czarnitzki and Stadtmann (2002) investigate the relationship between match uncertainty, calculated using betting odds, and live game attendance in the Bundesliga, which is the highest division in German football. Additionally, they include a variable for uncertainty in the outcome of the whole season through a measure that compares the points needed to win the Bundesliga that season to the points a certain team already has. Other variables include the reputation of a team, the number of supporter clubs, the size of the market, the form of the teams, weather conditions, the stage of the competition, and a dummy indicating if the match was broadcasted live on television. The measure for the reputation of a team is computed using the final rankings of a certain team over a period of 20 years. The authors argue that the attendance data is right censored, i.e. due to the capacity constraint of each stadium, 25% of the games were sold out and therefore true demand exceeded the maximum capacity of the stadium. This is the reason that Czarnitzki and Stadtmann (2002), and other authors mentioned later in this section, use a tobit model. Their results show coefficients that would support H1a. However, these coefficients were not statistically different from zero. The same applies to their uncertainty of outcome measures on the seasonal level. Therefore, Czarnitzki and Stadtmann (2002) conclude that the reputation of a team and the stage of competition are more important factors in explaining attendance.

Forrest and Simmons (2002) employ a similar approach as Peel and Thomas (1988; 1992) with regards to the model specification. However, they have concerns about the ability of the betting market to set betting odds unbiasedly. The authors therefore look into possible biases in betting odds. One bias that is backed by previous research is the favourite-longshot bias (Cain et al., 2000). The favourite-longshot bias states that "favourites win more often than the subjective market probabilities (set by the bookmakers) imply, and longshots less often" (Cain et al., 2000). Furthermore, Forrest and Simmons (2002) correct the betting odds for differences in mean attendances of the two teams playing, since the authors suspect that bookmakers vary their betting odds according to the relative level of support for the two teams. After correcting these biases they find evidence in favour of H3, using data from the English Football league excluding the English Premier League. Contrary to estimating a model for each division, fixed effects for each team are used to capture heterogeneity between the teams. Instead of the conventional functional form of the winning probability for the home team, Forrest and Simmons (2002) use the ratio between the probability of winning of the home team and the probability of winning for the away team as the uncertainty measure. With this uncertainty measure the classic UOH is supported in 97.8% of the football games the authors analysed. Forrest et al. (2005) also find evidence in favour of H3 with similar data and methodology.

The study by Falter et al. (2008) has its focus primarily on the effect of a World Cup victory of the national football team on the attendance demand in that country's domestic football league in the period after the World Cup victory. This research is somewhat different in terms of the main research question, but in their attendance demand model a measure for match uncertainty is included, which makes this study worth mentioning. Instead of using betting odds, Falter et al. (2008) use a self-computed measure of outcome uncertainty. It takes into account the points of the home team, points of the away team, and the home advantage. Their model explaining attendance includes variables of the World Cup victory effect, outcome uncertainty, stage of competition, quality of both teams and the match, team fixed effects, weather variables and a variable indicating if the match was broadcasted on live television. Their measure for outcome uncertainty proves to be insignificant in the OLS estimation for the French dataset that is used.

Madalozzo and Berber Villar (2009) also investigate the effect of outcome uncertainty on match attendance without the use of betting odds. The authors estimate fixed and random effects panel data models using data from the highest division in Brazil including variables on the quality of both teams and the match, cost of attending the match, stage of the competition and outcome uncertainty. Madalozzo and Berber Villar (2009) define outcome uncertainty in four variables; the difference in rankings between the home team and the away team, a probability measure for being the league leader, a probability measure of going to the Libertadores Cup[2], and a probability measure of leaving the rankings associated with relegation. For the uncertainty measures, the probability of being the league leader and probability of leaving the relegation zone were significant. Also the stage of competition had a significant influence on attendance. The approach by Madalozzo and Berber Villar (2009) use for their uncertainty measures is different compared to the model that is developed in this thesis. However, it is interesting to note that for instance the probability of being the league leader significantly increases game attendance.

The first cited study that looks into attendance demand in arguably the most popular league in the world (Curley & Roeder, 2016), the English Premier League, is a paper by Buraimo and Simmons (2008). Using betting odds as the basis for the uncertainty measure they estimate a tobit model, with

[2] The Latin American Football Cup, comparable to the UEFA Champions League.

comparable variables as the studies cited above. The results from this study are in line with H1a. A year later Buraimo and Simmons (2009) focused on another acclaimed football league, namely the Spanish Primera Division. For this study they estimate with similar variables a Prais-Winsten panel regression model, in which error terms are correlated across panels, i.e. home teams. Their results for stadium attendance are similar to their 2008 paper and support H1a. Interestingly, Buraimo and Simmons (2009) also look at the relationship between outcome uncertainty and the television viewer ratings for a particular match. They find evidence in support of the classic UOH with regards to television viewers.

In response to studies using OLS or tobit estimations, Benz et al. (2009) investigate demand for game attendance in the German Bundesliga with a distinctive approach. They employ a quantile regression model to overcome the assumption that average effects of the regressors apply to the whole distribution of demand. An advantage of quantile regression is that the importance of influence factors on attendance varies with the level of this attendance, which might be useful if for instance the home supporter's utility of their team winning is increasing in the number of home supporters. In this case, with quantile regression the probability of the home team winning has a greater influence for larger quantiles of attendance. Their data includes variables on the quality of both teams, quality of the match, economic factors and weather variables. Furthermore, they use a variety of measures for outcome uncertainty. These outcome uncertainty measures include the difference in rankings between the home team and the away team, a measure that compares the points per game for both teams and corrects for home advantage, and probabilities based on betting odds. Their results suggest evidence in support of H3. However, the coefficients for the uncertainty of outcome measures were only significant on a 10% significance level. Additionally, Benz et al. (2009) conclude that uncertainty of outcome is only a 'second-order' influence factor and that the reputation of a team is statistically more important in explaining attendance demand. The latter is also what Czarnitzki and Stadtmann (2002) found in their study of the Bundesliga.

Buraimo (2014) elaborated on his earlier research (Buraimo & Simmons, 2008; Buraimo & Simmons, 2009) by examining the five highest football divisions in England[3]. Similarly to Buraimo and Simmons (2008), the author estimates a tobit model for the English Premier League, since 47% of the games in this league are sold out in the dataset he uses. For other English leagues the author uses a panel data model with fixed effects. With the commonly used variable types each of the five leagues is analysed separately. Buraimo (2014) finds in each league results that confirm the reasoning of H1a.

[3] This includes the English Premier League, the English Football League Championship, the English Football League One, the English Football League Two, and the National League.

Reilly (2015) investigates the relationship between outcome uncertainty and live game attendance in the highest division in Ireland. Irish football is relatively small in terms of attendance compared to the leagues discussed above; in 2014 the average attendance was 1552.6 (Reilly, 2015). The author estimates, with conventional variables, different statistical models. The first notable model is comparable to the models that the other cited articles employ. In this fixed effects panel data model a linear and quadratic parameter is used for the probability of winning of the home team. This model confirms H1a. However, Reilly (2015) argues that the use of splines provides a more clear representation of the data. He uses two splines for the probability of winning of the home team in order to recreate the U-shaped relationship between probability of winning and match attendance as hypothesized by H1a. Of these two splines, only the first one is statistically significant. This implies that a negative relation exists between the probability of winning of the home team in the range of 0.0769 and 0.2471, and match attendance. Outside of this range of the win probability no significant influence on attendance could be found. In other words, matches where the home team has a probability of winning smaller than 0.25, there is less attendance the larger the probability of winning. The author concludes that the latter model dominates the quadratic model in terms of goodness-of-fit. Following the paper by Reilly (2015) is the paper by Jena and Reilly (2016), which examines the even smaller[4] second division in Ireland. The authors now choose to focus on a quadratic function of the probability of winning for the home team and use the conventional variables. Here they find evidence in support of the classic UOH (H3).

The last cited paper is a study by Martins and Cró (2016), which explores the highest Portuguese division. Next to the conventional variables, the authors use a variable for uncertainty of outcome on the seasonal level. Also, they argue that their variable for television broadcasted matches is subject to endogeneity, because the most attractive matches are chosen to be broadcasted. Therefore, Martins and Cró (2016) estimate a two-stage tobit model, using a time indicator as an instrumental variable for the television broadcast variable. The results obtained from this model support H1a.

3.1 Conclusions from the literature

Clearly, there is no consensus on the proposed relationship between match uncertainty and attendance in football. Coates et al. (2014) argue that this could be due to the fact that the specifications of the variables differ in these various papers. Furthermore, as is seen from the model developed in this thesis, the coefficients for the parameters of the probability of the home team winning reflect both the consumption utility (U^W and U^L) and the marginal utility of a win or a loss compared to the reference point (α and β). Cultural differences in the consumption utility might be

[4] Mean attendance in 2013-2015 was 464.23 (Jena & Reilly, 2016).

an important factor in the cited research. Also, consumption utility of sports might not be constant over time, since interest in sports varies over time, e.g. Falter et al. (2008) show that demand for football in a country rises if the national team of that country performs well. Consequently, it might be the case that studies from other countries and time periods are not externally valid and therefore it is interesting to conduct a study looking specifically at the Dutch Eredivisie with recent data.

4. Data & Methods

4.1 Data

The Dutch Eredivisie is the highest division in professional football in the Netherlands and comprises 18 teams. Within one season all teams play against each other twice; one home game and one away game. Points are awarded based on the outcome of the game. In case of one team winning, the winning team receives 3 points and the losing team receives no points. When two teams draw, both receive 1 point. The league rankings are based on the total amount of points each team has gathered throughout the season and at the end of the season the ranking of a team has certain consequences. Rewards for the highest ranking teams include winning the Eredivisie as champion or qualifying for (playoffs for) next season's European competitions. Negative consequences face the lower ranking teams, namely the risk of relegation to a lower division.

Attendance Data

Attendance data was collected from each club individually. One criterium for clubs to be considered was if their team played in the Eredivisie for most of the recent years. Clubs were contacted and asked if they were willing to participate. FC Groningen, Heracles and PSV were willing to participate with this research. Table 2 provides an overview of the attendance data that is used in this thesis. As can be seen from table 2, the number of observations in this dataset differs between clubs. The reason for this is that some clubs were not able to generate more data. The data gathered by clubs include total tickets sold for each home game, total season tickets sold that season and the maximum capacity of the stadium. Note that this data includes the total tickets sold and not total attendance on the day of the match. Data on the latter is not readily available.

Probabilities and betting odds

The probability of the home team winning is calculated using betting odds. The use of betting odds provides a market based approach to evaluate the probabilities of different outcomes. Betting odds data is collected from OddsPortal.com, where betting odds from the most used online bookmakers are collected. OddsPortal.com reports the mean betting odds collected from these bookmakers. The

Club		Mean	Minimum	Maximum	Observations
FC Groningen	Total tickets	20844.80	17818	22505	113
	Season tickets	13659.06	11442	15935	113
	Single tickets	7185.74	5510	10870	113
	Stadium capacity	22546	22546	22546	113
	Sold out	0.45	0	1	113
Heracles	Total tickets	9490.30	7985	12400	76
	Season tickets	5746.20	4600	7203	76
	Single tickets	3744.11	2795	5984	76
	Stadium capacity	9913.16	8500	12080	76
	Sold out	0.74	0	1	76
PSV	Total tickets	32940.96	29540	35031	202
	Season tickets	28042.93	25800	29157	202
	Single tickets	4898.03	3243	8847	202
	Stadium capacity	35000	35000	35000	202
	Sold out	0.42	0	1	202
Total	Total tickets	24886.95	7985	35031	391
	Season tickets	19552.06	4600	29157	391
	Single tickets	5334.89	2795	10870	391
	Stadium capacity	26524.55	8500	35000	391
	Sold out	0.49	0	1	391

Table 2 Descriptive statistics of attendance data. Games are sold out if the total tickets sold exceed the adjusted maximum capacity of the stadium. More information on the adjusted maximum capacity can be found in the methods section.

average amount of bookmakers for each game is 38.79 with a standard deviation of 12.33. Online bookmakers often adjust their betting odds in the days, hours or even minutes before a match. Betting odds reported on OddsPortal.com are the most recent betting odds. Usually this is not more than a few minutes before a game.

Betting odds incorporate important information on the competing teams, such as home advantage, missing players and current form. The assumption here is that betting markets are efficient and bookmakers have the incentive to be efficient in order to make profit. However, as mentioned by Forrest and Simmons (2002), some biases exist in betting markets. More recent research also confirms that backing strong favourites yields positive returns (Direr, 2013). This means that for games with a strong favourite the estimated probability of the favourite winning as constructed by the bookmaker is smaller than the actual objective probability. Due to data restrictions in this thesis it is not possible to correct the betting odds for possible biases using an equivalent approach as Forrest and Simmons (2002). Therefore in this thesis it is assumed that betting odds are indeed efficient. The probability of home team i winning the game against away team j is calculated using the formula

$$p_{ij} = \frac{x_{iw}^{-1}}{x_{iw}^{-1} + x_{id}^{-1} + x_{il}^{-1}} \tag{10}$$

where x_{iw}, x_{id} and x_{il} denote the betting odds[5] for a win, draw, or loss of the home team respectively.

Control variables

Data on the characteristics of both teams and the match are gathered from Voetbal.com. This data includes information about the standings of both teams prior to the match, the standings of both teams of last season, pre-match points of both teams, and pre-match goals scored and conceded for both teams. Pre-match standings for the first game of the season are none-existent, which is the reason for excluding the first game of the season from the dataset.

4.2 Methods

One of the aspects of football games for Dutch football clubs is the fact that a significant number of games sell out, as can be seen from table 2. This complicates the estimation of a demand function of attending a game, because 'real demand' exceeds the maximum capacity of a particular stadium. Therefore, sold out games are censored observations. One way of taking this into account is to estimate a tobit model. Tobit regressions take into account the fact that the dependent variable might be censored, which is the case in this dataset. Tobit models are also more clear in terms of the interpretation of the estimation, compared to methods such as the quantile regression used by Benz et al. (2009). The important thing to take into consideration here is the question of what exactly constitutes a sold out game. In the Dutch league away fans are commonly seated together in a small area of the stadium. These away fans have to be segregated from the home fans because of security reasons. In some cases this leads to a number of empty seats that provide a 'buffer' between rivalling fans, which causes the maximum capacity of the stadium to be lower. For the analysis, 5% of the maximum capacity is treated as empty seats that are due to security reasons, following the approach by Buraimo and Simmons (2008). The remaining net maximum capacity is used as the upper limit of the tobit model. This also solves the problem that in a few cases the reported total tickets sold exceeds the maximum stadium capacity by treating these few cases the same as other observations where the tickets sold exceeds the net maximum capacity. For sensitivity purposes a separate analysis is done to look at the differences in results when a different maximum capacity is used.

Furthermore, the data that is used in this thesis is panel data; the home teams are the panels.

[5] Betting odds are given in decimal format (EU format).

Therefore, it is most likely the case that the error terms are serially correlated because of unobserved time-constant home team characteristics. To account for this a panel data model can be estimated. This is a more efficient way of dealing with this sort of data as it exploits the correlation in the error terms. To do this the tobit model is extended to a random effects tobit model. A likelihood-ratio test confirmed that the random effects tobit model is superior to a pooled tobit model with the current dataset. The additional assumption that comes with a random effects tobit model as a panel data model is that the unobserved home team characteristics are not correlated with the independent variables. Madalozzo and Berber Villar (2009) show with their dataset on Brazilian football and a Hausman test that their random effects model is consistent and efficient, implying that unobserved characteristics are not correlated with the independent variables. This provides some evidence in favour of this assumption.

From the literature review certain variables were identified as possibly relevant in the analysis of attendance demand. Equation 9 serves as the basis for the econometric model that is estimated in this thesis. The model includes most relevant variables given that it provides a sufficient fit to the data according to the AIC and BIC, and the significance of the variables. More specifically, the model for home team i, away team j at time t with random error ε_{ijt} looks like:

$$\begin{aligned} \ln STS_{ijt} = \gamma p_{ijt}^2 + \theta p_{ijt} + \delta_1 RankHome_{ijt} + \delta_2 RankAway_{ijt} + \delta_3 Derby_{ijt} \\ + \delta_4 Renovate_{ijt} + \delta_5 Day_{ijt} + \delta_6 Month_{ijt} + \delta_7 Season_{ijt} \\ + \delta_8 Away_{ijt} + \varepsilon_{ijt} \end{aligned} \tag{11}$$

where:

$\ln STS_{ijt}$ = the natural logarithm of single tickets sold. A large number of consumers who attend a football match in the Netherlands are season ticket holders. In the dataset these season ticket holders are included in the total tickets. On average 74.51% of the total tickets sold are season ticket holders. This means that a significant part of the total tickets sold are fixed for the whole season, resulting in a rather large number of consumers to be unresponsive to any of the variables of interest. To counter this, in the analysis that follows the total tickets sold will be adjusted to reflect only the single tickets sold. This is done by subtracting the number of season tickets sold from the total tickets sold. The dependent variable for the random effects tobit model is therefore the logarithm of the single tickets sold. The transformation of the single tickets sold into a logarithm is done because the distribution of the single tickets sold is right skewed and taking the natural logarithm is a solution for this.

p_{ijt} = the probability of the home team winning, derived from betting odds with equation 10.

$RankHome_{ijt}$= the rank of the home team in the Eredivisie league table prior to the match. This variable captures the current form of the home team. Fans' interest might increase if their team is performing well in the current season. Also, the ranking of a team is a proxy for the relative quality of a team and fans might also be interested in seeing a quality team play.

$RankAway_{ijt}$= the rank of the away team in the Eredivisie league table prior to the match. This variable captures the current form of the away team. Home fans' interest might also be influenced by the form and quality of the visiting team.

$Derby_{ijt}$= a dummy variable indicating whether the match is considered a derby match. For the analysis in this thesis a derby is arbitrarily defined as a match where two teams of close geographical proximity play each other and winning the match has significant importance for fans of both clubs regardless of the quality/form of both teams. The latter is an added definition for this thesis. The reason for this is that including certain matches in the $Derby_{ijt}$ variable resulted in different levels of statistical significance. For instance, PSV plays other teams that are in the same region, such as NAC Breda or Willem II. However, adding these two matches to be considered for the $Derby_{ijt}$ variable resulted in an insignificant parameter. Therefore, these two matches for the final analysis are excluded, with the reasoning being that for PSV fans these matches are not more important compared to other matches they play against teams of the same quality. Two matches are ultimately considered as derby matches, namely FC Groningen against SC Heerenveen, the "derby of the north", and Heracles against FC Twente, the "Twentse[6] derby".

$Renovate_{ijt}$= a dummy variable that indicates in a given season if the stadium had a renovation. This variable is included to capture any effect that the excitement about a renovation of the stadium has on the home fans for the season following the renovation. In the dataset this constitutes to one particular season for Heracles, namely the 2015-2016 season.

Day_{ijt}= a vector of dummy variables indicating the day of the week. Controlling for the day of the week is important, because of different leisure or work-related activities are

[6] Twente is a region in the east of the Netherlands.

present on different days of the week, e.g. if a team plays midweek this might require more effort to attend a game for the home fan who has a full time job. Furthermore, the Eredivisie is organized such that from Tuesday until Saturday games are played in the evening, and on Sunday games are played in the afternoon[7]. Therefore controlling for the day of the week also controls for what part of the day the game is played. A separate variable for time of the day was experimented with, but this variable did not suit the dataset in terms of statistical significance.

$Month_{ijt}$= a vector of dummy variables representing the month in which the game is played. Fans' interest in attending a game might be prone to change over the course of a season, e.g. an increase in interest among fans can occur when the season is at its end, which means that the final rankings are close by. Also, including month dummies is a proxy for weather effects. Although the decision to buy a ticket, which is the data that is available for this thesis, is not influenced by the exact weather conditions on the day of the game, the decision to buy a ticket could very well be subject to the fans' expectations about these weather conditions. Including the month dummies captures these expectations about the weather.

$Season_{ijt}$= a vector of dummy variables indicating the season the game is played in. This control variable is a proxy for a number of things. Firstly, the economic environment in which a consumer operates is prone to changes over time. It controls for years in which an economic crisis was present and therefore acts as a proxy for changes in income. A decision of attending a game is likely influenced by the ability to afford the ticket. The assumption here is that these economic changes are the same for fans of all three clubs. Secondly, for the three clubs in the dataset ticket prices are constant over the course of a season. This means that a control variable for the season captures the changes in ticket prices. Finally, the seasonal dummies act as a proxy for changes in the way the Eredivisie is consumed through television broadcasting. The contracts for broadcasting rights are mostly fixed for each season and thus do not change during a season.

$Away_{ijt}$= a vector of dummy variables indicating the away team. This variable captures time invariant individual club effects. Certain visiting teams might attract more home fans to the games, e.g. when a visiting team has a high reputation higher attendance

[7] In the present dataset no games were played on Mondays.

Variable	Mean	Standard Deviation	Minimum	Maximum
$\ln STS$	8.5378	0.2968	7.9356	9.2938
p	0.5760	0.1911	0.1172	0.9078
$RankHome$	6.0358	4.7072	1	18
$RankAway$	9.5959	5.1251	1	18

Variable	Category dummy	Mean	Variable	Category dummy	Mean
$Derby$		0.0307	$Away$	AZ Alkmaar	0.0563
				Ajax	0.0563
$Renovate$		0.0435		Breda	0.0486
				Cambuur	0.0230
Day	Tuesday	0.0051		Den Bosch	0.0026
	Wednesday	0.0332		Den Haag	0.0563
	Thursday	0.0077		Dordrecht	0.0077
	Friday	0.0614		Excelsior	0.0358
	Saturday	0.4527		Feyenoord	0.0588
	Sunday	0.4399		G.A. Eagles	0.0205
				Graafschap	0.0256
$Month$	January	0.0716		Groningen	0.0435
	February	0.1407		Heerenveen	0.0588
	March	0.1125		Heracles	0.0460
	April	0.1049		Nijmegen	0.0486
	May	0.0460		PSV	0.0281
	August	0.0895		Roda	0.0537
	September	0.0972		Roosendaal	0.0051
	October	0.1151		Sparta Rotterdam	0.0179
	November	0.1023		Twente	0.0588
	December	0.1202		Utrecht	0.0614
				Venlo	0.0179
$Season$	2004-2005	0.0435		Vitesse	0.0588
	2005-2006	0.0435		Volendam	0.0026
	2006-2007	0.0409		Waalwijk	0.0307
	2007-2008	0.0435		Willem II	0.0486
	2008-2009	0.0435		Zwolle	0.0281
	2009-2010	0.0409			
	2010-2011	0.0844			
	2011-2012	0.0870			
	2012-2013	0.1279			
	2013-2014	0.1279			
	2014-2015	0.1228			
	2015-2016	0.1279			
	2016-2017	0.0665			

Table 3 Summary statistics of model variables

could be expected. The assumption with this variable is that these effects are constant over time.

Summary statistics for these variables are reported in table 3. Other variables were also considered, but were excluded since those variables did not suit the model in terms of significance and goodness-of-fit according to the AIC and BIC. Firstly, inspired by other literature a habit variable was added to control for a habit effect that might occur for home fans. The exact specification of this habit variable was constructed as the natural logarithm of single tickets sold from the previous home game.

Secondly, variations in the variable capturing the seasonal form of both teams were considered. These variations include the points per game average and a ratio for points over the maximum achievable points. The reason for experimenting with these variations is that these specifications provide a clearer picture of how well a team is doing, because the points are the underlying factor for determining the rankings. However, ultimately teams compete for a higher ranking and certain benefits that are awarded at the end of the season[8] are based on rankings, not the amount of points. This might be the reason why the absolute rankings are best suited for the model. Another variation that proved insignificant were the rankings of the previous season for both teams. Other variables comparing the number of points to the league leader or the last place in the league (including dummies for being the leader or last place) were found to be insignificant.

Finally, measures capturing the importance or quality of the match were considered. Two variations were experimented with, namely a measure that took the average ranking of both teams and a measure that captured the difference in points between two teams. Both these variations were insignificant.

5. Results

The estimated coefficients for the random effects tobit model are reported in table 4. Graphical examination of the residuals plotted against the fitted values showed that there was no problem with heteroscedasticity. The model shows quite a few interesting significant influences on the single tickets sold. Results, in terms of significance of the parameters, are robust when performing the same analysis with different levels of the adjusted maximum stadium capacity.

The variables of interest are p and p^2. Both variables have a significant coefficient at a 1% significance level. The signs of the coefficients represent clear support for H3, i.e. the 'classic' UOH. Figure 4 illustrates the relationship between the natural logarithm of single tickets sold and the probability of the home team winning. Generally, for the three clubs considered in this dataset more uncertainty means higher attendance in terms of the single tickets sold. The maximum $\ln STS_{ijt}$ is achieved at $p = 0.6003$, which indicates that the most single tickets are sold for games where the home team is slightly more likely to win than to not win. Note that for football, where games can end in a draw, a probability of winning equal to 60.03% does not mean that teams are more or less equal. Typically, when the probability of the home team winning is around 60%, the probability of a draw or the probability of the away team winning are around 23% and 17% respectively, according to the betting odds. This means that in this case the home team is roughly three times as likely to win compared to the winning chances of the away team. Examples of matches that have approximately the same winning probabilities are PSV-Vitesse in the 2014-2015 season, FC Groningen-Venlo in

[8] Such as the distribution of tv revenue or qualification for European competitions.

Variable	Category Dummy	Coefficient	Variable	Category Dummy	Coefficient
p		0.8933***	$Away$	AZ Alkmaar	‡
p^2		-0.7441***		Ajax	0.4781***
$RankHome$		-0.0014		Breda	0.0082
$RankAway$		-0.0063***		Cambuur	0.0536
$Derby$		0.1229**		Den Bosch	-0.0878
$Renovate$		0.1904***		Den Haag	-0.0590
Day	Tuesday	-0.0942		Dordrecht	0.2048**
	Wednesday	-0.0844*		Excelsior	-0.0470
	Thursday	-0.0266		Feyenoord	0.2143***
	Friday	0.0355		G.A. Eagles	0.0419
	Saturday	0.0325*		Graafschap	0.0665
	Sunday	‡		Groningen	0.0132
$Month$	January	‡		Heerenveen	0.0985**
	February	0.0370		Heracles	-0.0143
	March	0.1292***		Nijmegen	-0.0317
	April	0.2214***		PSV	0.2267***
	May	0.2707***		Roda	0.0510
	August	-0.0379		Roosendaal	-0.0622
	September	-0.0792**		Sparta Rotterdam	0.1122*
	October	0.0507		Twente	0.1263***
	November	0.0794**		Utrecht	0.0128
	December	0.0671**		Venlo	0.0693
$Season$	2004-2005	‡		Vitesse	0.0380
	2005-2006	-0.1134***		Volendam	0.0635
	2006-2007	-0.0990**		Waalwijk	0.0745
	2007-2008	-0.1305***		Willem II	0.0229
	2008-2009	-0.1190**		Zwolle	0.0758
	2009-2010	-0.0820*	Constant		8.2265***
	2010-2011	-0.0354			
	2011-2012	0.0042			
	2012-2013	-0.0352			
	2013-2014	0.0243			
	2014-2015	0.1148***			
	2015-2016	0.0684*			
	2016-2017	0.0833*			

*Table 4 Random effect tobit model coefficient estimates with dependent variable the natural logarithm of single tickets sold. *p<0.1, **p<0.05, ***p<0.01. ‡ indicates the reference category.*

the 2012-2013 season, and Heracles-Excelsior in the 2014-2015 season. As the literature in table 1 suggests, these results with evidence for the UOH contradict some of the papers, in particular the research by Coates et al. (2014), who developed the theoretical model that this thesis is based upon.

Other variables also influence single tickets sold significantly. For the current form of both teams, the rank of the away team has a significant negative coefficient. The negative sign is logically expected, since lower ranked teams most likely play less attractive football. Each additional rank for the away team results in a 0.63% decrease in single tickets sold[9], e.g. a 1.26% decrease is expected when playing an away team ranked number 11 instead of playing a team ranked number 9. In contrast, the rank of the home team as a proxy for current form has an insignificant coefficient.

The $Derby$ variable has a significant positive coefficient of 0.1229, indicating that when a derby

[9] All reported magnitude effects are ceteris paribus.

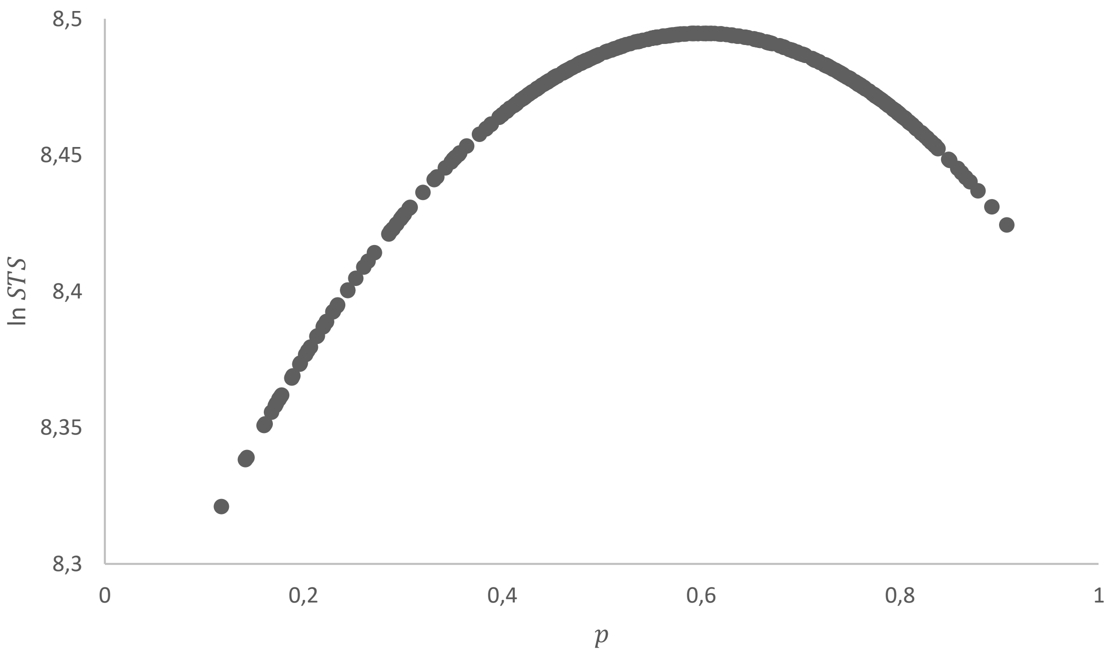

Figure 4 The relationship between the natural logarithm of single tickets sold and the probability of the home team winning as predicted by the estimated random effects tobit model.

match is played an increase of 12.29% is expected in single ticket sales. Once again it has to be noted that the exact definition of a derby match is arbitrary. FC Groningen-SC Heerenveen and Heracles-FC Twente are chosen as the derby matches in this dataset. As indicated in the methods section, other matches were also considered for model inclusion in the $Derby$ variable. These other matches seemed less important for home fans, as with these other matches included, e.g. PSV-Willem II, the $Derby$ variable had no significant influence. The conclusion from this is that for supporters of FC Groningen and Heracles the derby matches are more important in terms of prestige, compared to the other matches that could be considered a derby, such as regional derbies for PSV.

Another significant influence on the single tickets sold is the $Renovate$ variable. Since the variable is only applicable to one season and one club only, the interpretation is quite simple. In the 2015-2016 season single ticket sales were 19.04% higher for Heracles compared to other seasons. The fact that single ticket sales were higher compared to earlier seasons might be logical because the renovation itself mostly increased the stadium capacity. However, the variable also indicates that single ticket sales were higher in 2015-2016 compared to 2016-2017, which means that the attraction of a renovated stadium for the 2015-2016 season might have been a factor in single ticket sales.

Time variables also have a significant influence on single ticket sales. Firstly, two days of the week have a marginally significant coefficient ($p<0.1$), namely Wednesday and Saturday. Games on Wednesday are expected to have 8.44% less single tickets sold, compared to games in the reference category Sunday. Single ticket sales on Saturdays are 3.25% higher compared to the reference category Sunday. However, these day variables are only marginally significant. Secondly, the $Month$ variable has some categories that have a statistically significant influence in the model. The least single tickets sold are expected in September. Compared to the reference category January, games in September have 7.94% lower single ticket sales. For the other months single tickets sales are higher or do not differ from the reference category. November, December, March, April and March all have higher single ticket sales, respectively 7.94%, 6.71%, 12.92%, 22.14% and 27.07%. This clearly indicates that at the end of the season (March-May) single ticket sales increase, possibly because of the fact that games gain more importance as the season comes to an end. Finally, there are significant differences in single tickets sold comparing seasons. From season 2005-2006 until 2009-2010 there is a decline in single tickets sold compared to the reference season 2004-2005, with the coefficient for the 2009-2010 season being only marginally significant. The financial crisis around 2007/2008 could be a factor in this decline by influencing fans' ability to afford a single ticket. Another reason for lower single ticket sales in the years 2005-2008 could be the fact that the television rights for Eredivisie highlights and live broadcasting of Friday games were in the hands of commercial television channels in those years. This could have influenced fans' interest in football and the live broadcasting of Friday games on a 'free-to-view' channel could have been a substitute for attending a game in the stadium, since for the other seasons in the dataset Eredivisie games are only live broadcasted on paid channels. Seasons 2010-2011 until 2013-2014 did not have significantly different single ticket sales from the reference. From 2014-2015 onwards single tickets sold increased compared to the 2004-2005 season. The 2014-2015 season had the most single tickets sold in this dataset, with 11.48% more single tickets sold compared to the 2004-2005 season. The seasons after that also have higher single ticket sales, albeit with a marginal statistical significance. In 2015-2016 single tickets sales were 6.84% higher and in 2016-2017 8.33% higher, compared to 2004-2005.

For the away team dummies, some away teams had a significant coefficient. One away team clearly has the highest impact on single tickets sales compared to the other control variables, namely Ajax. If Ajax is the visiting team, single ticket sales increase with 47.81% compared to the reference away team AZ. PSV and Feyenoord, who together with Ajax comprise the 'traditional top three', also have a positive influence on single tickets sold when they are the visiting team, with increases in single ticket sales being 22.67% and 21.43% respectively, compared to the reference. Perhaps the most surprising result from the model is the effect that visiting team FC Dordrecht has on single ticket sales of the home team, with an increase of 20.48% compared to the reference category. The significant influence of visiting FC Dordrecht is most likely due to the fact that in the dataset there are only 3 matches where FC Dordrecht is the away team, all in the 2014-2015 season. In the 2014-2015 season FC Dordrecht did not perform well, and therefore home fans' expectations were most likely a dominant performance from the home team with a lot of goals. It was also the first time since 1995 that FC Dordrecht played in the Eredivisie. Other visiting teams that lead to an increase in single ticket sales are FC Twente, Sparta Rotterdam and SC Heerenveen, with 12.63%, 11.22% and 9.85% increases respectively, compared to the reference category.

6. Conclusion

With data on single tickets sold from 3 Dutch Eredivise clubs and betting odds data, this thesis provides evidence in favour of the 'classic' UOH, i.e. games with higher uncertainty yield higher live game attendance. The 'classic' UOH as it was first hypothesized was lacking a framework in which it could be analysed. The Coates et al. (2014) model provides a clear concept of how consumers decide to attend a game and together with the dataset in this thesis shows support for the UOH. For the UOH to be classified in the model the marginal utility of an unexpected win needs to exceed the marginal utility of an unexpected loss. The results show that the maximum single tickets sold is at a 60% winning probability for the home team. This is somewhat different from the hypothesized 55% winning probability by Rottenberg (1956), especially because in football three outcomes are possible, instead of two outcomes in most American sports. The analysis in this thesis shows that although the highest adjusted attendance is achieved at a fairly uncertain level of 60%, this maximum is still a case where the home team is a favourite to win, where the probability of the home team winning is roughly three times the probability of the away team winning. In American sports this would translate into a winning probability of 75% for the home team and 25% of the away team. Perhaps the best way to describe these results is a hybrid between the two opposite models. Clearly, uncertainty of outcome is optimal for the single ticket sales, but the maximum is achieved at a probability level that favours the home team and therefore loss aversion could likely still play a role in home fans' decisions.

The results obtained in this thesis are in line with conclusions from papers investigating English (Forrest & Simmons, 2002; Forrest et al., 2005), German (Benz et al., 2009) and Irish (Jena & Reilly, 2016) football leagues, but contradicts other papers from England (Peel & Thomas, 1988; Buraimo & Simmons, 2008; Buraimo, 2014), Spain (Buraimo & Simmons, 2009) and Portugal (Martins & Cró, 2016) that find evidence for the loss aversion model. As is said in the literature section, these differences can be caused by model specifications or differences in fans' attitude across countries, divisions within countries or time periods. Therefore, it is important to explore the relationship between uncertainty of outcome and live game attendance in the specific area of interest. This thesis only used data from 3 Eredivisie clubs. Future research looking at the Eredivisie would most likely benefit in terms of validity when considering more clubs.

If the objective of the KNVB, the governing body of Dutch football, was to increase live game attendance, the implications of this thesis are quite clear. In order to have the highest attendance, the probability of the home team winning has to be optimized according to figure 4, at least for the three analysed clubs. This is an argument for designing the league so that a large proportion of games follow this level of outcome uncertainty. One way in doing this is to make sure the distribution of player talent is more equal. This could be done by for instance more equally distributing tv revenues, which would result in a more equal distribution of available budgets for clubs. The English Premier League is a noteworthy example in which tv revenues are far more equally distributed (Total Sportek, 2017), compared to the Dutch Eredivisie (NOS, 2017).

Another way of optimizing the league is by reforming the structure of the league itself. One of the most striking proposals made to Dutch football clubs regarding a possible reform of the Eredivisie, is the proposal in which the number of participating Eredivisie clubs would be reduced from 18 to 16, and at the end of the season teams would compete in three separate groups for the championship, European qualification or relegation/promotion (NU, 2017). This is much like how the Belgian league is constructed. In turn this would mean more games with teams of comparable quality, compared to how the Eredivisie is currently organised. Only looking at the influence outcome uncertainty has on attendance, this thesis provides evidence in favour of this type of reform, simply because more games are played between clubs that are of comparable playing quality. Combining the comparable playing quality with the literature on home advantage (Pollard, 2008), there will be more matches that are close to the optimal level of outcome uncertainty of a 3 to 1 winning probability of the home team in this reformed league, compared to the current structure of the Eredivisie. However, outcome uncertainty is only part of what drives consumer demand for attending a game. Regardless of winning probability, the biggest influence on attendance is the visiting team, especially Ajax, Feyenoord and PSV. Out of the three clubs investigated in this thesis, PSV is most likely to benefit the most from the reforming of the league with a group stage at the end of the season. PSV is a club mostly competing for the championship title and is therefore most likely to encounter Ajax and Feyenoord (assuming both compete for the championship title as well) at the end of the season in the group playing for the championship. While these arguments favour the proposed reform over the current state of the Eredivisie, caution is needed when deciding for implementing the reform. The end of the season group stage would likely decrease the importance of games in the regular season. From the results in this thesis it can be shown that attendance is lower in the beginning months of the seasons, most likely due to the lack of importance of those games. With the reform it could very well be the case that this effect is extended to later months in the season, possibly countering the positive effects the reform has on stadium attendance. Also, with a separate model (not shown in this thesis) and the current dataset, there is a clear indication that in the league's current format total tickets sales (season tickets and single tickets) are significantly lower for playoff games at the end of the season, compared to regular league games. Further research is needed to examine the extent to which the fans are willing to attend these extra games at the end of the season. Data from the Belgian league shows an average 45% decrease in attendance for playoff games at the end of the 2016-2017 season in the group where 8 teams compete for one European qualification ticket, compared to regular league games (Voetbal.com, 2017). Judging from recent years, it is most likely that among the Dutch teams FC Groningen and Heracles would be in this middle group of 8 teams competing for that single European qualification ticket. Other playoff games in Belgium also seem to lack popularity among fans (Sporza, 2017). Depending on the club's ticket policy, there is also the issue that some fans have to pay extra ticket fees to attend the extra games at the end of the season. Affordability is therefore another important concern with a proposed reform of the Eredivisie.

Logically, there are more objectives that can have the priority for the KNVB. For instance, another important source of revenue for football clubs are television audiences (De Voogt, 2014). Buraimo & Simmons (2009) show that in Spain television audiences also follow the UOH, i.e. games with more uncertainty attract a larger television audience. While it is important to note that this would not necessarily hold for the Netherlands, this is more evidence in support of uncertainty in Dutch football games. For individual clubs the results provide a way in which they can predict the attendance of games they play. With knowledge on possible influences on live game attendance clubs are able to optimize their single ticket sales, e.g. through pricing or promotional strategies.

6.1 Limitations

The most important limitation of this thesis is the fact that the data consists of the total tickets sold, instead of actual attendance on the day of the match. While ticket sales are perhaps more important to football clubs, actual attendance is what is usually measured in the context of the model and literature. Furthermore, betting odds are prone to changes over time as the market constantly updates all relevant information and the betting odds that are used in this thesis were mostly set minutes before each match. It could therefore be the case that the probability of the home team winning at the time when a consumer decides to buy a ticket, would be substantially different from the probability of the home team winning hours or minutes before a game at a time where the consumer decides to actually attend the game. Since Eredivisie clubs face a certain no-show percentage (Verseput, 2012), the ticket sales would overreport the demand for the game when not taking this into account. If the no show-percentage would systematically differ between certain match characteristics, this would be an important factor to take into account. Another influence in overreporting of demand is the fact that clubs tend to provide free tickets to for instance businesses or young supporters, most likely for games that are not expected to sell out. Clubs do this to convey the impression of a full stadium, which can be useful in for instance sponsor revenues.

There is a large body of literature looking into home advantage, and specifically interesting in the context of this thesis are crowd effects (Pollard, 2008). It would be possible that there is presence of reverse causality, i.e. higher attendance influences the performance of a team and therefore the probability of winning or outcome uncertainty. For future research it might be interesting to look into this possible relationship.

In the relevant literature betting odds are often used as an instrument in calculating winning probabilities. However, as noted in the data section, betting odds can be subject to biases. According to Direr (2013) the probability of a strong favourite winning given by the bookmakers is actually lower than the actual ex post probability. This underestimation of winning probabilities for strong favourites could also lead to biases in the estimation of the model. Forrest and Simmons (2002) compare their results from an analysis including adjusted betting odds to account for possible biases and an analysis with betting odds as provided by the bookmaker. This showed that results from betting odds without adjustment provided less evidence in favour of the UOH, compared to results with adjusted betting odds. In identifying definitive evidence in favour of the UOH, it would be a necessary step to include an adjustment in the betting odds in future research.

7. References

Benz, M.-A., Brandes, L., & Franck, E. (2009). Do Soccer Associations Really Spend on a Good Thing? Empirical Evidence on Heterogeneity in the Consumer Response to Match Uncertainty of Outcome. *Contemporary Economic Policy Vol. 27, No. 3*, 216-235.

Buraimo, B. (2014). Spectator Demand and Attendances in English League Football. In J. Goddard, & P. Sloane, *Handbook on the Economics of Professional Football* (pp. 60-72). Cheltenham, UK: Edgar Elgar Publishing.

Buraimo, B., & Simmons, R. (2008). Do Sports Fans Really Value Uncertainty of Outcome? Evidence from the English Premier League. *IntertMtional Journal of Sport Finance Vol. 3, No. 3*, 146-155.

Buraimo, B., & Simmons, R. (2009). A tale of two audiences: Spectators, television viewers and outcome uncertainty in Spanish football. *Journal of Economics and Business Vol. 61, No. 4*, 326-338.

Cain, M., Law, D., & Peel, D. (2000). The favourite-longshot bias and market efficiency in UK football betting. *Scottish Journal of Political Economy Vol. 47 No. 1*, 25-36.

Card, D., & Dahl, G. (2011). Family Violence and Football: The Effect of Unexpected Emotional Cues on Violent Behavior. *Quarterly Journal of Economics Vol. 126, No. 1*, 103-143.

Coates, D., Humphreys, B. R., & Zhou, L. (2014). Reference-Dependent Preferences, Loss Aversion, and Live Game Attendance. *Economic Inquiry Vol. 52, No. 3*, 989-973.

Curley, J. P., & Roeder, O. (2016, March 17). *English soccer's mysterious worldwide popularity*. Retrieved March 2017, from Contexts: https://contexts.org/articles/english-soccers-mysterious-worldwide-popularity/

Czarnitzki, D., & Stadtmann, G. (2002). Uncertainty of Outcome Versus Reputation: Empirical Evidence for the First German Football Division. *Emperical Economics Vol. 27, No. 1*, 101-112.

De Voogt, S. (2014, August 8). *Zo verdienen de Eredivisieclubs dit seizoen miljoenen*. Retrieved May 2017, from NRC.nl: https://www.nrc.nl/nieuws/2014/08/08/dit-is-wat-alle-18-clubs-uit-de-eredivisie-komend-seizoen-verdienen-a1500527

Direr, A. (2013). Are betting markets efficient? Evidence from European Football Championships. *Applied Economics Vol. 45 No. 3*, 343-356.

Falter, J.-M., Pérignon, C., & Vercruysse, O. (2008). Impact of Overwhelming Joy on Consumer Demand: The Case of a Soccer World Cup Victory. *Journal of Sports Economics Vol. 9, No. 1*, 20-42.

Forrest, D., & Simmons, R. (2002). Outcome Uncertainty and Attendance Demand in Sport: The Case of English Soccer. *Journal of the Royal Statistical Society. Series D (The Statistician) Vol. 51, No. 2*, 229-241.

Forrest, D., Beaumont, J., Goddard, J., & Simmons, R. (2005). Home advantage and the debate about competitive balance in professional sports leagues. *Journal of Sports Sciences Vol. 23, No. 4*, 439-445.

Jena, F., & Reilly, B. (2016). Testing the uncertainty outcome hypothesis using data from second tier soccer in Ireland. *Applied Economics Letters Vol. 23, No. 18*, 1257-1260.

Kahneman, D., & Tversky, A. (1979). Prospect Theory: An Analysis of Decision under Risk. *Econometrica Vol. 47, No. 2.*, 263-292.

Köszegi, B., & Rabin, M. (2006). A Model of Reference-Dependent Preferences. *The Quarterly Journal of Economics Vol. 121, No. 4*, 1133-1165.

Madalozzo, R., & Berber Villar, R. (2009). Brazilian Football: What Brings Fans to the Game? *Journal of Sports Economics Vol. 10, No. 6*, 639-650.

Martins, A. M., & Cró, S. (2016). The Demand for Football in Portugal: New Insights on Outcome Uncertainty. *Journal of Sports Economics*, 1-25.

Neale, W. C. (1964). The Peculiar Economics of Professional Sports: A Contribution to the Theory of the Firm in Sporting Competition and in Market Competition. *The Quarterly Journal of Economics Vol. 78, No. 1*, 1-14.

NOS. (2017, May 13). *Laatste speeldag: de keiharde strijd om de tv-miljoenen*. Retrieved June 2017, from NOS.nl: http://nos.nl/artikel/2172916-laatste-speeldag-de-keiharde-strijd-om-de-tv-miljoenen.html

NU. (2017, May 17). *Eredivisie blijft bestaan uit 18 clubs na mislukt overleg over nieuwe opzet*. Retrieved June 2017, from NU.nl: http://www.nu.nl/voetbal/4700859/eredivisie-blijft-bestaan-18-clubs-mislukt-overleg-nieuwe-opzet.html

Peel, D., & Thomas, D. (1988). Outcome Uncertainty and the Demand for Football: an Analysis of Match Attendances in the English Football League. *Scottish Journal of Political Economy Vol. 35, No. 3*, 242-249.

Peel, D., & Thomas, D. (1992). The Demand for Football: Some Evidence on Outcome Uncertainty. *Empirical Economics Vol. 17, No. 2*, 323-331.

Pollard, R. (2008). Home Advantage in Football: A Current Review of an Unsolved Puzzle . *The Open Sports Sciences Journal Vol. 1*, 12-14.

Reilly, B. (2015). The Demand for League of Ireland Football. *The Economic and Social Review Vol. 46, No. 4*, 485-509.

Rottenberg, S. (1956). The Baseball Players' Labor Market. *Journal of Political Economy Vol. 64, No. 3*, 242-258.

Sporza. (2017, April 4). *Fans liggen nog niet wakker van play-offs: slechts 64.470 toeschouwers*. Retrieved June 2017, from Sporza.be: http://sporza.be/cm/sporza/voetbal/Jupiler_Pro_League/1.2942292#

Total Sportek. (2017, June 6). *Premier League Prize Money 2017 (Complete breakdown of entire tv money fund)*. Retrieved June 2017, from TotalSportek.com: http://www.totalsportek.com/money/premier-league-prize-money/

Verseput, S. (2012, February 10). *Bijna uitverkocht? Nee, een halfvolle Arena*. Retrieved May 2017, from NRC.nl: https://www.nrc.nl/nieuws/2012/02/10/bijna-uitverkocht-nee-een-halfvolle-arena-12164409-a1396029

Voetbal.com. (2017, May 17). *Play-offs Europa League niet populair in België*. Retrieved June 2017, from Voetbal.com: http://www.voetbal.com/news/_n2748475_/play-offs-europa-league-niet-populair-in-belgie/