SCHOOL OF
ECONOMICS AND
MANAGEMENT

LUND
UNIVERSITY

The Effect of Expected Outcome on Attendance

Evidence from Swedish Allsvenskan

by

Ismael Hallbäck and Johannes Reuter

May 2021

Master's Programme in Economics

Supervisor: Roel van Veldhuizen

Abstract

Live attendance at sports events is important for individual teams as well as for the broader community of a sport. The amount of uncertainty about the outcome of a game is considered a crucial determinant of game attendance. This study examines how fans' expectation about game outcome affects game attendance in Swedish Allsvenskan. The empirical analysis is guided by a behavioural model of the decision to attend sports events. Using a fixed effects estimator, we find that fans show a preference for games with little uncertainty. This can be explained by reference-dependent preferences and loss aversion.

Keywords: prospect theory, attendance demand, football, uncertainty of outcome hypothesis

Table of Contents

1 Introduction ... 1

2 Swedish Allsvenskan ... 3

3 Behavioural Model for the Decision to Attend Sports Events ... 4

3.1 Theoretical Background ... 4

3.2 The Model ... 5

3.3 Analysis of the Model ... 6

4 Game Attendance in Football ... 8

4.1 Attendance Demand in Football ... 8

4.2 Uncertainty of Outcome Hypothesis ... 9

5 Estimation Strategy ... 12

6 Data ... 15

6.1 Data Source ... 15

6.2 Variables ... 15

6.3 Sensitivity Analysis ... 19

7 Results ... 21

7.1 Raw Data ... 21

7.2 Main Analysis ... 22

7.3 Sensitivity Analysis and Robustness Checks ... 25

8 Discussion ... 28

9 Conclusion ... 30

References ... 31

Appendix A ... 34

Appendix B ... 35

1 Introduction

The audience is of great importance for sports events, both for the individual team and for the sport at large. For a team, it is vital for the matchday revenue that many people attend. For the community around the sport, a high attendance raises welfare. Fans are offered a package, which includes an audience that is chanting, booing, and creating an atmosphere in the stadium. High attendance is also important for external parties, such as TV broadcasters and sponsors. Thus, it is important to understand how game attendance is determined. Moreover, an essential feature of all sports events is that the outcome is not established beforehand. When the probabilities of the possible outcomes are known, fans' choice between attending and not attending a game is a decision made at risk. This makes fans' attendance decisions interesting to analyse from an economic point of view.

The *uncertainty of outcome hypothesis (UOH)* established by Simon Rottenberg (1956) is the first attempt to explain how attendance demand works. Rottenberg (1956) states '*That is to say, the "tighter" the competition, the larger the attendance*.' (Rottenberg, 1956, Footnote p.246). and '*uncertainty of outcome is necessary if the consumer is to be willing to pay admission to the game*' (Rottenberg, 1956, p.246). Neale (1964) continues on these lines and coins the term "Lewis-Schmeling paradox", saying that unlike a firm where the optimum is to hold monopoly power, Joe Lewis, the heavyweight champion in boxing could only make money because he had a good opponent in Max Schmeling. Although the statements provide a clear intuition, they have little connection with formal economic theory.

Recently, Coates, Humphreys, and Zhou (2014) have developed a behavioural model of the decision to attend sports events. They try to capture fans' actual thoughts and behaviour and not the choice a rational individual should make. The model builds on ideas from *Prospect Theory* by Kahneman and Tversky (1979), in particular the concept of *reference-dependence*, which means that utilities are evaluated relative to a reference point and not as final welfare. The model predicts the UOH, as well as several other explanations for attendance demand. For example, that fans show preferences opposite to the UOH. That is, fans prefer games with little uncertainty over games with a lot of uncertainty.

Empirical research covers many of the factors that influence game attendance, but the main focus is on the UOH. In the research, uncertainty is understood in terms of fans' expectations about the outcome. A game is uncertain if fans have expectations that are very close for two or more of the possible outcomes. However, the results of recent studies on the UOH are mixed and mainly cover US sports. Thus, there is a need for more research on attendance demand.

The purpose of this study is to contribute to the research by investigating football in Sweden, in particular the highest division for men, *Swedish Allsvenskan*. To the best of our knowledge, there is no published study on this topic. European football is interesting to focus on as it has a different institutional background than most US sports. The research question is *'what is the effect of the expected outcome on attendance for football games?'*

We answer the research question using betting data from *Svenska Spel* and attendance data from *Svenska Fotbollsförbundet*. The answer should be guided by economic theory. So, the first step is to model fans' decision process and link it to observed game attendance. The next step is to get a broader understanding of attendance demand and to develop an estimation strategy that has a clear connection to economic theory. The theoretical foundation, as well as the estimation strategy, are based on Coates, Humphreys, and Zhou (2014). In practice, we test whether game attendance is influenced by the winning probability of the team on who's stadium the game takes place. Moreover, we test if there is a positive or a negative relationship, alternatively, a convex or a concave relationship. Our results could be important to further understand how game attendance is determined. Especially, they can provide an explanation that is firmly grounded in economic theory about decision-making under risk.

Our paper is structured as follows; the next section describes the institutional background of Swedish Allsvenskan. Section 3 focuses on theory and presents a model of the decision to attend sports events. Section 4 contains a literature review of the research on game attendance. Section 5 and 6 describe the estimation strategy and the data set, respectively. Section 7 reports the results, sensitivity analysis, and robustness checks. Discussion is given in section 8 and section 9 concludes.

2 Swedish Allsvenskan

This section describes the institutional background of Swedish Allsvenskan. A concise description helps to give a clear and exact answer to the research question ‘what is the effect of the expected outcome on attendance for football games?’

Swedish football follows a structure that is common across Europe. The organisation Svenska Fotbollsförbudent (SvFF) is an elected body and one of its main responsibilities is the national league. This is a tiered league with promotion and relegation between divisions. One special feature of Swedish football is that the ownership of teams follows the so-called ’51-per cent rule’. This is the requirement that at least 51% of the shares are owned by members and not by commercial interests (SVT, 22 May 2013).

Allsvenskan is the highest division in the Swedish football league for men. It consists of 16 teams and is a double round-robin tournament. This means that the teams play against each other twice; one game at home and one game away (Allsvenskan, 2021). For winning a game they receive 3 points, for a draw they receive 1 point, and if they lose, they receive 0 points (Svenska fotbollsförbundet, 2021, p.9). When the season ends, the team with the most points wins the league. The two teams with the fewest points are relegated to the second-highest division, while the team finishing third to last plays a qualification match against the third-best team in the second-highest division (Svenska fotbollsförbundet, 2021, p.11).

The structure of Allsvenskan results in a very open league. Teams have no exclusive right to geographic areas. For example, there are three teams from Stockholm that play in the season 2021; AIK, Djurgårdens IF, and Hammarby IF. On the other hand, there are no geographic areas that are guaranteed to be represented in the league. For example, until 2017 there was no team from Uppsala, which is Sweden’s fourth-largest urban area.

In conclusion, Swedish Allsvenskan is characterised by its administrative structure and the format of the competition. This type of structure seems to be favourable for a strong supporter culture. Moreover, there is the prerequisite for high competition across the league; the best teams aim to win, and the worst teams fight to stay in.

3 Behavioural Model for the Decision to Attend Sports Events

This section summarises the insights derived from the behavioural model developed by Coates, Humphreys, and Zhou (2014). The model is highly relevant as it makes predictions related to the research question 'what is the effect of the expected outcome on attendance for football games?' Before explaining the theoretical results, the section gives a brief description of the model and the economic theory on which it builds.

3.1 Theoretical Background

In the influential paper "*Prospect Theory: An Analysis of Decision under Risk*" (1979), Kahneman and Tversky criticise *Expected Utility Theory*. Although based on rational choice, Expected Utility Theory is not always consistent with real-world behaviour. As a solution, Kahneman and Tversky (1979) introduce an alternative model which they call *Prospect Theory*. One important component of *Prospect Theory* is *reference dependence*. This means that outcomes are not evaluated as final welfare, but instead measured relative to a reference point and then coded as gains or losses relative to this reference point. The model also introduces the notion of *loss-aversion*, meaning that losses feel worse than same-sized gains.

By focusing on reference-dependent preferences, Koszegi and Rabin (2006) create a more generally applicable model based on *Prospect Theory*. Firstly, the reference point is determined within the model. Unlike earlier work, where the reference point is some status-quo, the reference point is based on rational expectations of future outcomes. Secondly, the model incorporates decision-making under certainty. Individuals evaluate outcomes in two ways and overall utility consists of two separate parts. *Consumption utility* is the outcome-based utility that a rational individual would base their decision on. *Gain-loss utility* is the utility due to the sensation of gain or loss relative to the reference point. In the context of sports, one of the most important applications of this model is by Card and Dahl (2011), in which they connect domestic violence with the outcome of National Football League games. The same framework is later used by Coates, Humphreys, and Zhou (2014) to analyse attendance demand for Major League Baseball.

3.2 The Model

The analysis in this study is based on the behavioural model of fans' decision to attend sports events, developed by Coates, Humphreys, and Zhou (2014). This model is useful for analysing game attendance in sporting events as it incorporates behavioural economic concepts and has a connection to the observed attendance. The model is developed for baseball, which differs from football as games cannot end in a draw. However, the model can still be applied to football if the expectation of a draw does not influence fans' decision to attend a game.

The individuals in the model are sports fans, and they evaluate the risky option of attending a game that their team plays in the home stadium. A game has two outcomes; either the home team wins or loses, which are denoted y = 1 and y = 0, respectively. The expectation of the game outcome is the *probability* that the home team wins: E[y = 1] = *p*. This is common knowledge among fans. In addition, the utility of individuals is based on reference-dependent preferences. One part of fans' utility is *consumption utility*, which is based on the outcome of a game. If the home team wins, the individual gets $U(1) = U^W$ and if the home team loses, the individual gets $U(0) = U^L$. The assumption is that individuals have a *home win preference*, that is $U^W > U^L$. The second part of fans' utility is *gain-loss utility*, which is measured as the deviation from the reference point. As expectations about the outcome are rational, the reference point is *p*. Further, the positive constants α and β are the weights that individuals attach to a positive respective negative gain-loss utility. The resulting utility function has two parts, one for home team wins and one for home team losses:

$$\begin{cases} u(p) = U^w + \alpha(y - p), if\ y = 1 \\ u(p) = U^L + \beta(y - p), if\ y = 0 \end{cases}$$

$$\begin{cases} u(p) = U^w + \alpha(1 - p), if\ win \\ u(p) = U^L - \beta \cdot p, if\ loss \end{cases} \tag{1}$$

As can be seen in equation 1, the individual gets a positive gain-loss utility if the home team wins. This part of overall utility is increasing in the weight attached to a positive gain-loss utility, and negatively related to the probability of the home team winning. That is, the more unlikely it is that the home team wins, the higher is the positive gain-loss utility. If the home team loses, the individual gets a negative gain-loss utility. This part of overall utility is decreasing in the weight attached to a negative gain-loss utility, and negatively related to the probability of the home team winning. That is, the more likely it is that the home team wins, the higher is the disutility of a negative gain-loss utility.

Furthermore, the assumption is that individuals are risk-neutral expected utility maximisers. So, the prospect of attending a game is evaluated by calculating expected utility:

$$E[U] = p[U^W + \alpha(1-p)] + (1-p)[U^L + \beta(0-p)]$$
$$E[U] = (\beta-\alpha)p^2 + [(U^W + U^L) - (\beta-\alpha)]p + U^L \qquad (2)$$

Equation 2 relates a fan's expected utility of attending a game to *p*, the probability that the home team wins. Expected utility is a quadratic function of *p*. Moreover, a fan attends a game when the expected utility is greater than *v*, the *reservation utility* from not attending a game (Coates, Humphreys, & Zhou, 2014).

3.3 Analysis of the Model

Coates, Humphreys, and Zhou (2014) make several predictions for the relationship between expected outcome and game attendance. The predictions are derived from equation 2 and depend on the values of the model parameters. This section focuses on the predictions that are most relevant for this study. Firstly, the model covers the situation when fans have preferences without reference dependence. This holds when $\alpha = \beta = 0$. In this case, the expected utility of attending a game is increasing in *p*, the probability that the home team wins the game. Thus, game attendance is predicted to be increasing in *p* (Coates, Humphreys, & Zhou, 2014).

Furthermore, the model also predicts the uncertainty of outcome hypothesis (UOH), which has a prominent role in the literature. The hypothesis states that game attendance is increasing in the amount of uncertainty about the outcome. One way to compare the amount of uncertainty between games is to measure how close the probability of a home team win is to the probability of a home team loss. The outcome of a game is completely certain when $p = 0$ or 1, and the highest uncertainty is at $p = 0.5$. The prediction holds when $(\beta - \alpha) < 0$ and $U^W - U^L < \alpha - \beta$ (Coates, Humphreys, & Zhou, 2014). That is, the weight attached to a positive gain-loss utility is greater than the weight attached to a negative gain-loss utility. In addition, fans' preference for a positive gain-loss utility dominates the positive marginal utility of a home win. This means graphically that fans' expected utility function is concave with respect to *p* and has a maximum at $p = 0.5$ or just above (Coates, Humphreys, & Zhou, 2014).

The model also gives another prediction supported by empirical evidence; the opposite to the UOH, which means that game attendance is decreasing in the amount of uncertainty about the outcome. This prediction holds when $(\beta - \alpha) > 0$ and $0 \leq U^W - U^L < \beta - \alpha$ (Coates, Humphreys, & Zhou, 2014). The first condition states that the weight attached to a negative gain-loss utility is greater than the weight attached to a positive gain-loss utility. So, a negative deviation from the reference point results in a greater disutility than the utility of a positive deviation of the same size. This is simply loss-aversion. The second condition states that an individual's aversion for a negative gain-loss utility dominates the positive marginal utility of a home win. The graphical interpretation of the prediction is that fans' expected utility function is convex with respect to p, with a maximum at $p = 0$ or 1 and a minimum between these two points (Coates, Humphreys, & Zhou, 2014).

Coates, Humphreys, and Zhou (2012) relate the behavioural model to observed game attendance, which makes the model useful in answering the research question of this study. They suggest that game attendance is equal to the number of fans with an expected utility greater than their reservation utility. They also assume all fans have the same expected utility function and the reservation utility is identically distributed. So, game attendance is defined the following way:

$$Attendance = PotentialAttendance \cdot P(v \leq E[U]|p) \tag{3}$$

Potential attendance is the number of individuals who would consider attending a game independent of the probability of a home team win, and $P(v \leq E[U]|p)$ is the probability that an individual has an expected utility of attending greater than the reservation utility, conditional on the expected outcome p (Coates, Humphreys, & Zhou, 2012).

In summary, this section focuses on a model of fans' decision to attend sports events. The model builds on prospect theory and one important feature is that fans' utility consists of two parts. One for normal consumption utility and another for reference-dependent gain-loss utility. In addition, this section explains how to analyse the decision process of an expected utility maximising fan. The model gives several predictions for the relationship between the expected outcome of a game and attendance, of which the UOH and its opposite are the most relevant for this study. Finally, the chapter briefly mentions how the model relates to observed game attendance. This connection is important for the empirical analysis and is therefore developed further in the section that covers the estimation strategy.

4 Game Attendance in Football

This section examines how existing literature contributes to answering the research question 'what is the effect of the expected outcome on attendance for football games?' The literature guides the empirical analysis and provides support for the theoretical predictions. The first part puts the research question into context. It discusses determinants of attendance demand in football. The second part focuses on the uncertainty of outcome hypothesis. It gives a summary of recent research on the topic, with an emphasis on European football.

4.1 Attendance Demand in Football

A football game is a product that consists of more factors than just the expected outcome. This part aims to show how expected outcome and the related concept of uncertainty of outcome fits into a broader understanding of attendance demand in football. The insights are important for the empirical analysis. This part describes the factors that influence the demand for football games, which can roughly be organised into three categories: *fan interest*, *quality of the view*, and *opportunity cost of attending* (Borland & MacDonald, 2003).

Football fans attend a game because of interest, and the expected outcome is one of several aspects of fan interest. The theory section shows that the expected outcome can influence fan interest in several ways, for example through the amount of uncertainty about the game outcome. Fan interest also depends on the *quality of the game*. This is defined as the skill shown by players (Borland & MacDonald, 2003) and can be measured in many ways. There is evidence that game attendance is influenced by the rankings of the competing teams (Reilly, 2015), the recent performance of the home team and the visiting team (Cox, 2018), and the expected number of total goals in the game (Peel & Thomas, 1992). Another determinant of fan interest is the *significance of a game*. Peel and Thomas (1992) write that fan interest might be influenced by the significance of a game in the context of the league. They also suggest that the significance of a game depends on the combination of teams that play. Evidence in line with this idea is the impact a geographical derby has on game attendance (Reilly, 2015). Fan interest might also depend on the team brand (Pawlowski &

Anders, 2012) and famous players (Schreyer, Schmidt, & Torgler, 2018). Finally, fans are interested simply because they support the playing team (Peel & Thomas, 1992).

Additional factors that influence attendance demand are less connected to the expected outcome of a game. The condition of the stadium at which the game takes place might be important for the quality of the view (Peel & Thomas, 1992). So is the timing of the game. There is evidence that the time of the kick-off, the day of the week, and the weather conditions affect game attendance (Reilly, 2015). Moreover, the opportunity cost of attending depends on the alternative activities on which fans can spend time and money. Some factors related to the opportunity cost of attending are alternative seasonal attractions (Forrest & Simmons, 2002), games in rival leagues or other sports, and the possibility to see the game on tv (Reilly, 2015).

Overall, the literature suggests that many factors of a diverse nature influence attendance demand in football. The factors that are most important for this study are those which are related to fan interest. These factors are very likely to correlate with the expected outcome. Factors related to the quality of view and opportunity cost of attending are less likely to relate to the expected outcome but are common as control variables.

4.2 Uncertainty of Outcome Hypothesis

Ever since Rottenberg (1956), there have been several studies evaluating the existence of the uncertainty of outcome hypothesis (UOH). Traditionally, the UOH has been agreed upon by the literature, as can be seen by Szymanski’s (2003) summary of earlier studies. However, Szymanski (2003, p. 1182) questions whether the conclusion holds for European football. This part presents recent research and discusses insights that are important for empirical analysis.

There are roughly three levels of uncertainty connected to games. The research question covers the expected outcome of a specific game. Thus, the UOH always refers to the concept of game uncertainty in this study. Another concept is seasonal uncertainty. The uncertainty of a specific team winning or losing a season is one example of how to define this (Peel & Thomas, 1992). There is also league uncertainty. This can be defined as the predictability of team performance over several seasons (Peel & Thomas, 1992).

Previous studies find significant effects supporting the UOH in the US national football league, NFL (Paul, Wachsman, & Winbach, 2010), US national basket association, NBA (Mills & Fort, 2014), US college football (Eckard, 2017), Swiss Super League (Baranzini, Ramirez, & Weber, 2008), the Irish league (Reilly, 2015), and in the Championship, second tier in England (Lahvicka, 2013). Other studies have found significant effects disproving the UOH in the Premier League (Cox, 2018), Portuguese first division (Martins & Cró, 2018), Serie A (Di Domizio & Caruso, 2015), and the highest football league in England, Spain, Italy, and Germany (Serrano, Garcia-Bernal, Fernandez-Olmos, 2015). Although there are exceptions, some overall patterns can be seen. The UOH seems to hold in American sports and the lower leagues of European football. Conversely, the UOH seems to fail in the highest leagues of European football.

The observation about attendance demand might be connected to differences in the institutional background of the leagues studied. For example, in the US, the league owns the teams, which are seen as franchises. In addition, the major US sports leagues are not open but closed, meaning that no team gets relegated after the season is over. It is also common to use a playoff system in these leagues. This means that at the end of the season the top teams of a league play a single-elimination tournament to establish the winner. Games before the playoff might matter less than early games in European football. Moreover, we have two ideas that can potentially explain the mechanisms at work. Firstly, if the quality of a game is high enough, this might outweigh any influence uncertainty has on fan interest. Secondly, if the supporter culture in a league is weak, fan interest might mainly be driven by the suspense of a game and not by the desire to see a specific team winning. However, we can only speculate about the exact connection between institutional background and the characteristics of attendance demand.

Furthermore, we found several papers that are of importance for our empirical analysis. The study by Coates, Humphreys, and Zhou (2014) on major league baseball, MLB, uses a fixed effects estimator and finds a significant result that contradicts the UOH. They find that fans seem to have reference-dependent preferences and loss-aversion. In Buraimo and Simmons' (2009) study of the highest division in Spain, they use the probability of a home team win and its squared values and find significant results; again, not in line with the UOH. They also included away team variables for Barcelona and Real Madrid arguing that the biggest clubs in the country will attract more fans, regardless of the home team win probability.

In the research, there is variation in the measures used for the expected outcome. Besters, van Ours, and van Tuijl’s (2019) study of the highest division in the Netherlands use both the probability of the home team winning, its squared value, and a new measurement called ‘match expectation’, which is the home teams expected points and its squared value. They argue that this captures the probability of a draw in a more correct way than using the home team win probability. Both measurements find evidence against the UOH. In addition, Benz, Brandes, and Franck (2009) use a variable for uncertainty of outcome developed by Roy (2004) and finds evidence against UOH in the highest football league in Germany. Jespersen and Pedersen (2018) focus their paper on four major European football leagues, the highest league in England, Germany, Spain, and Italy. They reject the UOH in all four leagues. They use an uncertainty of outcome variable developed by Theil (1967). Sacheti, Gregory-Smith, and Paton (2014) mention that results can be sensitive to different measures of uncertainty. So, alternative measures for expected outcome and variables for uncertainty of outcome are useful to include in a sensitivity analysis.

In conclusion, the literature indicates that the expected outcome of a football game is relevant for fans’ decision to attend. The literature seems to support the predicted relationships between expected outcome and game attendance, which suggests that the framework outlined in the theory section is useful for analysing attendance demand. For sports in general, there is mixed support for the UOH, and the UOH does not seem to hold for the highest leagues of European football. There is some evidence of the opposite prediction, which requires reference dependence and loss-aversion. The institutional structure of European football is a potential explanation for the observed pattern. In addition, previous research suggests that results can be sensitive to the measure of outcome uncertainty.

5 Estimation Strategy

This section explains how to estimate the effect of the expected game outcome on attendance. The goal is to relate observed game attendance to the behavioural model. First, it presents a regression model that combines economic theory and statistical assumptions, as suggested by Coates, Humphreys, and Zhou (2014). Then, we explain the advantage of using a fixed effects estimator instead of a naive OLS estimator. We also explain how to control for other sources of bias. Finally, we discuss the remaining limitations of the estimation strategy.

To analyse observed game attendance, Coates, Humphreys, and Zhou (2012) combine the results of the theory section with a statistical model. They use equation 3 to formulate a hypothesis about the relationship between the expected outcome and game attendance:

$$\ln(\mathrm{ATTENDANCE}_{ijt}) = \lambda + \theta p_{ijt} + \gamma p_{ijt}^2 + X_{ijt}\mu + \varepsilon_{ijt} \text{ [1]}$$

The derivation of specification [1] is found in Appendix A. The dependent variable is the natural logarithm of the *attendance* for a game in which team *i* plays at home against team *j* at season *t*. In the economic model potential attendance is assumed to have an exponential functional form, which is the reason that the natural logarithm is taken on attendance (Coates, Humphreys, & Zhou, 2012). The main explanatory variable is the home team win probability, p_{ijt}, which has a quadratic relationship to the outcome. The parameters for p_{ijt}, and p^2_{ijt} are $\theta = [(U^W - U^L) - (\beta - \alpha)]/(\bar{v} - \underline{v})$ and $\gamma = (\beta - \alpha)/(\bar{v} - \underline{v})$, respectively (Coates, Humphreys, & Zhou, 2014). The parameters are directly connected with the parameters of the behavioural model. This relates the econometric specification to the preferences and predictions in the theory section. For example, the estimate supports the uncertainty of outcome hypothesis if $\gamma < 0$ and $\theta > 0$. On the other hand, it supports the opposite where fans have reference-dependent preferences and loss aversion if $\gamma > 0$ and $\theta < 0$. If $\gamma = 0$ the result suggests that fans do not have reference-dependent preferences (Coates, Humphreys, & Zhou, 2014)**.** The remaining explanatory variables are contained in X_{ijt}. These variables represent other factors influencing game attendance and are related to potential attendance. For example, characteristics of the home team and away team, and dummy variables for the day of the week and month of the year (Coates, Humphreys, & Zhou, 2014). Finally, ε_{ijt} is the error term that captures all unobserved home team and away team characteristics that influence the potential attendance of a game (Coates, Humphreys, & Zhou, 2014).

Ideally, the error term ε_{ijt} is uncorrelated with all unobservable home team and away team characteristics that influence attendance (Wooldridge, 2016, p. 76). However, this cannot be ensured when using observational data, which is the only type of data that can realistically be obtained. So, it is problematic to run a naive OLS estimation on specification [1]. The estimation is likely to be biased due to omitted variables. In the case of football, the win probability of teams is likely to relate to attendance through underlying factors. This can be thought of as the quality of the team, which covers factors such as the players in the team, team management, and the financial resources spent. The quality of a team is likely to influences fan interest as well as the expected outcome. This holds for home games, as well as away games, and is many times difficult to observe. To overcome the potential bias due to the quality of teams, Coates, Humphreys, & Zhou (2014) suggest that fixed-effects dummy variables are included in the econometric specification:

$$\ln(\text{ATTENDANCE}_{ijt}) = \lambda + \theta p_{ijt} + \gamma p_{ijt}^2 + X_{ijt}\mu + D_i + D_j + D_t + \psi_{ijt} \quad [2]$$

Specification [2] contains dummy variables for each team for home games, D_i, as well as away games, D_j. and for each season, D_t. The regression parameters can be estimated using an OLS estimator, that is *the ordinary least squares dummy variable method* (Verbeek, 2008, p. 345). The dummy variables control for all time-invariant factors related to the home team and the away team which affect potential attendance. (Wooldridge, 2016, p. 412). Note that the factors do not necessarily need to be time-invariant, but as good as time-invariant over the period of interest (Wooldridge, 2016, p. 413). Examples of factors are team quality, the number of people living in the home team's city, and the number of loyal home team fans. In addition, the season dummies control for general time trends that affect potential attendance for all games in a specific season (Wooldridge, 2016, p. 412). One example of a trend factor is the general interest and popularity of Swedish football at a certain point in time.

More control variables can be included in the regression to reduce the influence of other sources of bias. A dummy for team combinations, D_{ij}, controls for time-constant factors related to games where team *i* meets team *j* in team i's home stadium. One example is the effect of derbies. Note that D_{ij} and D_{ji} are likely to capture the same effects. A team-specific seasonal effect, D_{it}, controls for factors affecting potential attendance that are constant for all games that team *i* plays at home in season *t*. One example is increased interest in team *i*'s home games because the team has been promoted from a lower division league. Note that D_{it} and D_{jt}, are likely to be, if not the same, at least very similar.

Although we aim to control for several sources of bias, it is still necessary to be cautious when interpreting the results. For the estimation to be unbiased it is necessary that the remaining error term, ψ_{ijt}, is uncorrelated with the main variable of interest, p_{ijs}, in all seasons t and s (Wooldridge, 2016, p. 432). This is a very strong assumption that is unlikely to hold in the context of football game attendance. The process of how attendance and the probability of a game outcome are determined is complex. There are many factors, and it is close to impossible to control for all. Especially factors that vary during the season and influence the expected outcome as well as attendance through fan interest, the quality of the view, and the opportunity cost of attendance. In addition, the data set limits the number of control variables that can be included in the regression. However, the types of controls mentioned so far are likely to be the most important ones. So, even if the estimation is biased, it is likely to hold qualitatively.

Furthermore, there are additional limitations with the estimation approach, but we do not believe that these have an important influence on our study. The data panel is unbalanced, and the reason for attrition is frequent losses which lead to a team's exclusion from the league. This implies a sample selection based on team performance. However, as the frequency of losses is unlikely to be based on the dependent variable, attendance, the OLS estimator is unlikely to be biased due to attrition (Wooldridge, 2016, p. 440). Further, the efficiency of the fixed-effects estimation requires error terms that are serially uncorrelated across time and homoscedastic (Wooldridge, 2016, p. 436). As part of the robustness check, the homoscedasticity assumption is tested.

In summary, this section presents a regression model for the relationship between expected outcome and game attendance, which is based on the behavioural model for fans' attendance decision. The estimated regression parameters indicate if the relationship is in line with the theoretical predictions and the corresponding fan preferences, in particular the UOH and its opposite. To overcome omitted variable bias, a fixed effects estimator is used, which controls for time-invariant factors related to the home team or the away team. Other sources of bias important to control for are factors related to the combination of teams that play and factors that affect a team in a specific season. Although the estimation strategy cannot eliminate all sources of bias, these are unlikely to influence the results qualitatively.

6 Data

This section describes the data set used to estimate the effect of the expected outcome on attendance in Swedish football games. The data collection is guided by the estimation strategy and the literature review. First, we introduce the data source. Next, we define each variable and give descriptive statistics. Finally, we discuss alternatives to the main explanatory variable to be used in a sensitivity analysis.

6.1 Data Source

The empirical analysis is based on an unbalanced panel of football games over three seasons of Swedish Allsvenskan, 2017-2019. The main source for the data set is Svenska Fotbollsförbundet, the association that organises Allsvenskan. Besides information on the outcome variable of this study, attendance, they also publish the following information about each game: the home team, the away team, the stadium, and the game result (Svenska fotbollsförbundet, 2021). This information was downloaded and transformed into a data set. In each season 240 games were played, and the full sample contains 720 observations. In total 21 teams are included in the data set, and a table of the teams and the seasons they take part in is found in the appendix, table B1. Furthermore, the main explanatory variable, the probability of a home team win, is based on data from *Svenska Spel*. This is a state-owned gambling company in Sweden. The data set contains information on their closing odds for each game for the same period as the main data set. The data is not publicly available and was provided on request. Our final data set is derived by merging the two data sets. Further, the closing odds for the first two rounds in 2017 are missing. So, we exclude these data points, and the final data set contains 704 observations.

6.2 Variables

Table 1 gives an overview of all variables. The definitions for the outcome variable and the main variable of interest originate from the estimation strategy. The estimation strategy also indicates what type of control variables should be included. The exact definition of the control variables comes from the literature review.

Table 1 Descriptive Statistics

	Mean	Std.Dev.	Min	Max	Obs
HOME_i			1.00	21.00	720
AWAY_j			1.00	21.00	720
ATTENDANCE	8799.25	7166.06	202	50128	704
lnATTENDANCE	8.78	0.81	5.31	10.82	704
PROB_H	0.45	0.18	0.06	0.84	704
PROB_H^2	0.23	0.16	0.00	0.70	704
PROB_D	0.25	0.04	0.11	0.33	704
PROB_A	0.31	0.16	0.04	0.80	704
Control Variables					
PROMOTED	0.19	0.39	0.00	1.00	720
DERBY	0.04	0.19	0.00	1.00	720
FIRST	0.02	0.15	0.00	1.00	720
AVGGOAL_H	1.31	0.48	0.00	5.00	712
AVGGOAL_A	1.34	0.46	0.00	3.00	712
AVGCONCED_H	1.33	0.48	0.00	3.00	712
AVGCONCED_A	1.31	0.49	0.00	5.00	712
AVGPOINTS_H	1.34	0.56	0.00	3.00	712
AVGPOINTS_A	1.39	0.56	0.00	3.00	712
HABIT_H	8.74	0.82	6.55	10.06	720
HABIT_A	8.74	0.82	6.55	10.06	720
RANK	0.54	0.50	0.00	1.00	720
SEASON			2017	2019	720
MONTH			3.00	11.00	720
WEEKEND	0.62	0.49	0.00	1.00	720
Sensitivity Analysis					
BRAND_A	0.31	0.46	0.00	1.00	720
THEIL	0.99	0.11	0.54	1.10	704
ROY	0.58	0.11	0.20	0.69	704
POINTS_H	1.58	0.51	0.32	2.63	704
POINTS_H^2	2.77	1.60	0.10	6.90	704

The outcome variable *lnATTENDANCE* is based on data on game attendance, which corresponds to the variable *ATTENDANCE*. *ATTENDANCE* is defined as the number of individuals that attend a game, but Svenska *Fotbollsförbundet* does not specify how this variable is recorded. As can be seen in Table 1, *ATTENDANCE* takes values between 202 and 50 128, with an average value of 8799. Graphs B1 and B2 in Appendix B show that most games have an attendance around the mean, but there are some games with a deviation far to the right. Graph B3 in appendix B shows that for away games, the distribution of attendance on a team level is similar to the overall distribution, but for home games there seems to be less variation in attendance within teams and more variation across teams.

As stadiums have a maximum capacity, there is a possibility that *ATTENDANCE* is right-censored. Of the 704 games, 27 (3,84%) are sold out using the notion that sold out is when attendance is 95% of the capacity. Alternatively, 11 (1,56%) games are sold out if using the notion that sold out is when attendance is equal or higher than the maximum capacity. Among the sold-out games, attendance is sometimes above the maximum capacity, which we think suggests that the maximum capacity only takes into account a seated audience and not a standing audience. So, stadiums might have a higher capacity than stated. We think it unlikely that right censoring influences the results, because the number of sold-out games is small, and stadiums might have higher capacity than stated.

The definition of the dependent variable, *lnATTENDANCE*, has implications for the interpretation of estimated coefficients. Because of the regression model, the dependent variable is defined as the natural logarithm of *ATTENDANCE*. For functions of the form $\log(y) = \beta_1 + \beta_1 x$, we can approximate a change in y due to a small change in x using the following formula, $\%\Delta y \approx (100\ \beta_1)\Delta x$ (Wooldridge, 2016, p. 638). The coefficients of the explanatory variables approximate the proportionate change in the dependent variable due to a change in an independent variable. So, the coefficients multiplied by 100 should be interpreted as the percentage change in attendance due to a unit change in the variable in question (Wooldridge, 2016, p. 172).

The main explanatory variable is $PROB_H$, which is the probability that the home team wins. The probability is computed from the closing odds recorded by 'Svenska Spel'. The odds are expressed as the pay-out ratio of a winning bet. We want to compute the probability that team *i* wins a game without the odds-setters markup influencing the result. So, we take the inverse of the odds and weigh it by the inverse of the sum of the inverses of the odds of all three possible outcomes, home team win, draw, and home team loss (Kuyppers, 2000):

$$PROB_i = \frac{1}{Odds_i} * \frac{1}{\sum_{i=1}^{3} \frac{1}{Odds_i}}$$

Table 1 shows that the probability that the home team wins takes values between 0.06 and 0.84, with an average value of 0.45. In Appendix B graph B4 shows that the overall distribution of home win probability seems normally distributed around the average. Graph B5 in the appendix indicates that there are differences in the distribution of home win probability across teams.

The regression model contains $PROB^2_H$, which is the squared value of $PROB_H$. This has implications for the interpretation of the results. For quadratic functions of the form y = β_1 + β_1x + β_2x^2, we can approximate a change in y for a small change in x using the following formula, $\Delta y \approx (\beta_1 + 2\beta_2 x)\Delta x$ (Wooldridge, 2016, p. 635). So, the coefficients of both $PROB_H$ and $PROB^2_H$ should be taken into consideration when interpreting the relationship between $PROB_H$ and attendance. For example, if the probability of a home team win increases by 0.1 and the original probability is p, then the percentage change in attendance is approximately $100*(\beta_1 + 2\beta_2 p)\cdot 0.1$.

Furthermore, the characteristics of the probabilities of the game outcomes impact the interpretation of $PROB_H$. Table 1 shows that the values of the probability that the home team loses are relatively low compared to the values of the probability that the home team wins. This is also the case for $PROB_A$, the probability that the away team wins, which takes values between 0.04 and 0.8. The two probabilities do not add to one due to the presence of the probability of a draw. As can be seen in table 1, $PROB_D$, the probability of a draw, takes values between 0.11 and 0.33, with an average value of 0.25. Graph B6 in the appendix shows that the probability of a draw is the highest just below $PROB_H$ = 0.4 and decreases to both sides of this value. Graph B7 shows that a game is equal when $PROB_H$ is just below 0.4, and not at 0.5. In addition, Graph B7 also shows that for a given home win probability, there is little variation in the probability that the away team wins. This also implies that there is little variation in the probability of a draw, which supports the notion that the expectation of a game ending in a draw has a limited impact on the attendance decision of fans.

The control variables originate from the estimation strategy and the literature review. Dummy variables for the home teams and away teams are included in the regression to capture time-invariant fixed effects of the home team- and away team market. So, for each of the 21 teams, we create dummies $HOME_i$ and $AWAY_j$. The dummy is present when team *i* plays a game at their home stadium or when team *j* plays a game at the opponent's home stadium.

The regression also contains time-related dummy variables. $SEASON_t$ is a dummy for each season and captures factors that affect attendance in a specific season. *FIRST* is a dummy for the first game of the season, which tends to attract more attendance than normally. Other time-related dummies capture the viewing quality of the game. *MONTH* is a set of dummies, one for each month of the year that games are played, March to October, but leaving out April. *WEEKEND* is a dummy that is present if the game takes place on a weekend.

Additional control variables are related to fan interest. To capture the effect of the quality of a game we have $AVGGOALS_H$ and $AVGGOALS_A$, which are the average goals scored by the home team and the away team, respectively, before the game and within the same season. $AVGCONCED_H$ and $AVGCONCED_A$ are the average goals conceded by the home team and the away team, respectively, before the game and within the same season. $AVGPOINTS_H$ and $AVGPOINTS_A$ are the average points collected by the home team and the away team, respectively, before the game and within the same season. Recent performance can also increase fan interest. So, we include *PROMOTED*, a dummy to capture the team-specific seasonal effect of newly promoted teams. The *DERBY* dummy accounts for derby games within seasons, these are when both teams are from Stockholm, both teams are from Gothenburg or the “Skåne-derby” between Malmö FF and Helsingborg IF. We include the dummy *RANK* to account for the attractiveness of a club. *RANK* is made so that the five teams with the most points in the previous season get a 1, otherwise 0. Lastly, we want to control for the support of a team. Thus, we include $HABIT_H$ and $HABIT_A$, which captures the effect of persistent attendance. The variable is defined as the team’s average attendance in the previous season.

6.3 Sensitivity Analysis

We perform sensitivity analysis to see how robust our results are. Firstly, we want to check if the use of a variable for uncertainty of outcome influences the main result. One of the earliest variables for uncertainty of outcome was developed by Theil (1967):

$$THEIL = \sum_{i=1}^{3} p_i \, log \, (\frac{1}{p_i})$$

Where p_i, i = 1,2,3 denotes the probability for a home team win, a draw, and an away team win. The smaller the difference is between the probabilities, the larger is the value of *THEIL*.

Roy (2004) criticizes Theil’s measurement because it includes the probability of a draw. The argument is that if winning probabilities are equal for both home and away teams, then a change in draw probability would decrease the value for *THEIL* (Benz, Brandes, & Franck, 2009). For instance, if we assume home win probability and away win probability each decreases from 0.35 to 0.3, and draw probability increases from 0.3 to 0.4, then the probabilities would not change relative to each other. However, *THEIL* would decrease. Roy (2004) suggests an adjusted measure, *ROY*, which just uses the probabilities of a home team win and an away team win. Unlike the *THEIL*, the value for *ROY* stays the same when the probability of a draw increases but the relationship between home win and away win is constant. The adjusted measure is defined in the following way

$$ROY = \sum_{k=1}^{2} \frac{P_k}{P_1 + P_2} \log\left(\frac{P_1 + P_2}{P_k}\right)$$

Where p_k, $k = 1,2$ denotes the probability for a home team win and an away team win. Again, the smaller the difference is between the probabilities, the larger is the value of *ROY*. For both these measurements, it must be that the coefficients are positive to support uncertainty of outcome hypothesis. One reason for not using these parameters in our main analysis is the fact that they would not be able to capture reference dependence. Hence, the results can only be interpreted as evidence towards whether fans prefer certain or uncertain outcomes.

We also include $POINTS_H$ and $POINTS^2_H$ which are the expected points for a home team i during a home match. According to Besters, van Ours, and van Tuijl (2019), this expression for expected outcome takes into account the possibility of a draw in a better way. Since a win gives 3 points, a draw gives 1 point and a loss gives 0 points, the variables are computed as

$$POINTS_H = 3 * (PROB_H) + 1 * (PROB_D) + 0 * (PROB_A)$$

A dummy, $BRAND_A$, captures the effect of the away team's brand image. According to Pawlowski and Anders (2012), this is another and fairer assessment of attractiveness, compared to *RANK*. They find evidence that the inclusion of this variable makes the expected outcome variable insignificant in the German Bundesliga. From a survey study on sports brands (Sport & Affärer, 2021), five clubs, AIK, Hammarby IF, Djurgården IF, IFK Göteborg, and Malmö FF have a substantially larger perceived sporting success and perceived brand image than the rest of the clubs, hence if these clubs are the away team, they will receive a 1 and otherwise 0.

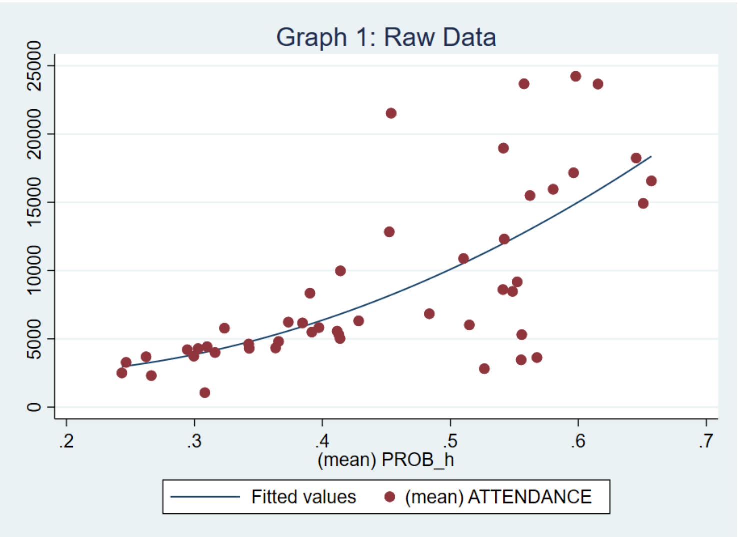
In conclusion, the empirical analysis builds on a data set that merges betting odds from 'Svenska Spel' with information on football game attendance for three seasons of Swedish Allsvenskan. The estimation strategy suggests that the outcome variable is the natural logarithm of attendance and that the main explanatory variable is the probability of the home team winning. We also include control variables related to time and fan interest in the data set. Lastly, two different variables for uncertainty of outcome, as well as an alternative variable for expected outcome based on expected points, can be used in a sensitivity analysis.

7 Results

This section presents the results of the empirical analysis. The estimated relationship between expected outcome and attendance indicates if fans behave in line with the theoretical predictions and suggest what type of preferences they have. First, we investigate the raw data. Then we describe the main analysis. Finally, present the result of the sensitivity analysis and the robustness check.

7.1 Raw Data

Before we do a detailed investigation of the relationship between attendance and the expected outcome, we examine the raw data for an initial understanding of the connection. Graph B8 in the appendix shows the overall correlation between a home team win probability and attendance. There seems to be a weakly positive correlation. On the other hand, graph B9 in the appendix shows the correlation between home team win probability and attendance on a team level, and in most cases, there seems to be a negative correlation. These two observations might be reconciled by graph 1, which shows the average attendance and the average probability of a home team win for each team and each season.



The graph indicates a positive correlation between the average probability of home team win and the average attendance. Although attendance might not increase in home team win probability on team level, it seems that teams with higher average home team win probability also have higher average attendance. In addition, the functional form of the fitted line suggests a quadratic relationship. For a more reliable analysis we run a fixed-effects regression.

7.2 Main Analysis

Table 2 presents our main results. All the columns are regressions of the natural logarithm of attendance and the main variable of interest $PROB_H$, the probability of a home team win. Furthermore, all regressions include home, away, and time fixed effects.

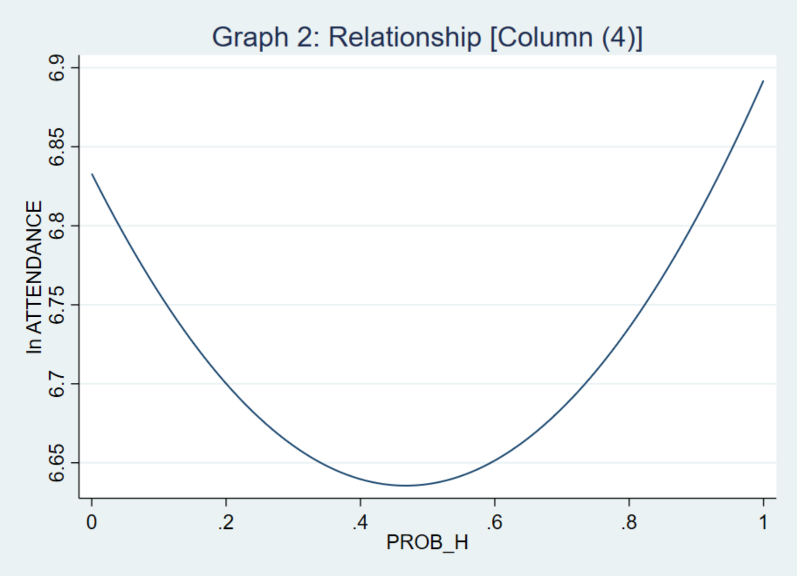
Table 2 Main Results

	ln Attendance			
	(1)	(2)	(3)	(4)
$PROB_H$	-0.8553**	-0.7633**	-0.7550*	-0.8454**
	(0.3738)	(0.3848)	(0.3855)	(0.4023)
$PROB_H^2$	0.8376**	0.8202**	0.8183**	0.9044**
	(0.3678)	(0.3666)	(0.3670)	(0.3829)
WEEKEND	0.0420*	0.0410*	0.0415*	0.0428*
	(0.0229)	(0.0228)	(0.0228)	(0.0229)
PROMOTED	0.1794	0.1744	0.1651	0.2069
	(0.1423)	(0.3362)	(0.3369)	(0.3411)
DERBY	0.2530***	0.2602***	0.2608***	0.2643***
	(0.0610)	(0.0607)	(0.0609)	(0.0610)
FIRST	0.3003***	0.2975***	0.2973***	0.2974***
	(0.0876)	(0.0877)	(0.0878)	(0.0878)
$AVGGOAL_H$		0.0146	-0.0244	-0.0246
		(0.0352)	(0.0631)	(0.0631)
$AVGGOAL_A$		0.0283	0.0453	0.0517
		(0.0311)	(0.0565)	(0.0571)
$AVGCONCED_H$		-0.1042**	-0.0733	-0.0767
		(0.0417)	(0.0615)	(0.0616)
$AVGCONCED_A$		-0.0830**	-0.1012*	-0.1026*
		(0.0366)	(0.0570)	(0.0570)
$HABIT_H$		0.1564	0.1266	0.2680
		(0.9413)	(0.9434)	(0.9605)
$HABIT_A$		0.0586	0.0622	0.0693
		(0.0711)	(0.0722)	(0.0728)
$AVGPOINTS_H$			0.0538	0.0525
			(0.0719)	(0.0720)
$AVGPOINTS_A$			-0.0242	-0.0308
			(0.0635)	(0.0641)
RANK				-0.0282
				(0.0356)
Home-FE	YES	YES	YES	YES
Away-FE	YES	YES	YES	YES
Time-FE	YES	YES	YES	YES
R-squared	0.9041	0.9064	0.9065	0.9066
Observations	704	704	704	704

Notes: Standard Errors in parentheses. *, **, *** denote statistical significance at the 0.10, 0.05, 0.01 level.

In column (1), we include control variables for promoted teams, derby, and the first game of the season. In column (2) we add control variables for average goals scored and average goals conceded for home team and away team, up until the game and within the same season. In column (3), we include control variables for average points collected for home team and away team, up until the game and within the same season. In column (4) we include a control variable to account for the attractiveness of the teams.

From table 2, our main interest is the result for the expected outcome, $PROB_H$ and $PROB^2_H$. The results are significant, at the 0.05 level and the 0.10 level. For all columns, the sign for the variable $PROB_H$ is negative and the sign for the variable $PROB^2_H$ is positive. This suggests that the relationship between expected outcome and attendance is convex, see graph 2. The coefficients do not change too much between the columns; $PROB_H$ is between -0.75 and -0.85, $PROB^2_H$ is between 0.90 and 0.81. In column (1) the minimum value for attendance is when the home win probability lies at 0.51, from the other columns it lies around 0.46.



Recall, that for the uncertainty of outcome hypothesis, UOH, to hold, the relationship between the probability of home win and *lnATTENDANCE* must be concave. Meaning that $PROB_H$ should have a positive sign while $PROB^2_H$ should have a negative sign. On the other hand, fan behaviour shows a preference for outcomes that are certain if the relationship between the probability of home win and *lnATTENDANCE* is convex. This relationship holds if $PROB_H$ has a negative sign and $PROB^2_H$ has a positive sign. Thus, the main analysis seems to contradict the UOH and instead suggests the opposite. This can be seen from the effect of the expected outcome on attendance, as indicated by the coefficients of $PROB_H$ and $PROB^2{}_H$. From column (4) we have that the variable $PROB_H$ has the coefficient -0.8454 which is significant at the 0.05 level, and $PROB^2{}_H$ has the coefficient 0.9044, also significant at the 0.05 level. So, if the probability of a home win increases from 0.1 to 0.4, then the proportionate change in attendance is approximately (-0.8454 +2·0.9044·0.1)·0.3 = -0.1994, or a decrease by 19.9%. Further, if the probability of a home win decreases from 0.8 to 0.5, then the proportionate change in attendance is approximately (-0.8454 +2·0.9044·0.8) · (-0.3) = -0.1805, or a decrease by 18.1%.

FIRST is positive and significant in all columns, this implies that the first game of the season attracts more attendance. The coefficient in column (4) is 0.2974, at a 0.01 significance level, meaning that the first game of the season attracts on average 29.74% more attendance compared to other games. Also, *DERBY* is positive and significant in all columns, implying that matches against geographic rivals attract more attendance. The coefficient in column (4) is 0.2643, at a 0.01 significance level, meaning that a derby will on average increase attendance by 26.43% compared to other games. Further, attendees seem to prefer games on weekends rather than weekdays, or more likely, attendees have more time during weekends to go to a match. The coefficient for *WEEKEND* in column (4) is 0.0428, at a significance level of 0.10, this should be interpreted as weekends will on average increase attendance by 4.28% compared to weekdays. In column (4), $AVGCONCED_A$ seems to be significant at a 0.10 level with a coefficient of -0.1026. It is a rather surprising result that the more goals the away team concedes, the fewer people attended. It could relate to the downward sloping part of the convex relation, meaning that fans have an interest in seeing good teams play.

Finally note that when using the ordinary least squares dummy variable method, R-squared is almost always very high. This is to be expected as a lot of the variation in the data set can be explained by the dummies (Wooldridge, 2016, p. 436). So, by itself, the value of R-squared does not contain a lot of information about the explanatory variables.

7.3 Sensitivity Analysis and Robustness Checks

Table 3 presents the sensitivity analysis. First, we include the variable $BRAND_A$ in the original regression. $BRAND_A$ captures the effect of the away team brand image on attendance. This is to control that brand image is not the main explanatory variable instead of the expected outcome. Note that the fixed effect of away teams already accounts for brand image, so there is perfect multicollinearity between a dummy for away brand and the fixed effect of away team dummies. Therefore, away team fixed effects are not included in column (1). We also try two different measurements of uncertainty of outcome, to control that the results are not sensitive to the choice of the main explanatory variable. These are *THEIL* and *ROY*. Finally, we try an alternative measure of the expected outcome, $POINTS_H$ and $POINTS^2_H$.

In column (1), $BRAND_A$ is significant and $PROB_H$ and $PROB^2_H$, are still significant with the same signs as our main results. This contradicts Pawlowski & Anders's (2012) result that brand image is the main explanatory variable instead of the expected outcome. In column (2) and column (3), *THEIL* and *ROY* are significant, and the coefficients are negative. This suggests a positive relationship between attendance and certain outcomes. In column (4), $POINTS_H$ and $POINTS^2_H$ are significant and follow our main results. The attendance minimum is at 1.57 points. Overall, the sensitivity analysis is in line with the main result; the expected outcome is a good explanatory variable and fans prefer certain outcomes.

Finally, we want to make sure that our results are robust even if the error terms are heteroscedastic. So, Table 4 shows the main regressions run with robust standard errors, which ensures that the test statistics are appropriate (Verbeek, 2008, p. 88). The results from these regressions are similar to our main results, which indicates that our main results are qualitatively robust.

Table 3 Sensitivity Analysis

	ln Attendance			
	(1)	(2)	(3)	(4)
$PROB_H$	-0.6340*			
	(0.3735)			
$PROB_H^2$	0.8400**			
	(0.3652)			
THEIL		-0.2558**		
		(0.1243)		
ROY			-0.2311*	
			(0.1243)	
$POINTS_H$				-0.3071**
				(0.1489)
$POINTS_H^2$				0.0967**
				(0.0448)
$BRAND_A$	0.2421***			
	(0.0360)			
WEEKEND	0.0378*	0.0421*	0.0415*	0.0423*
	(0.0229)	(0.0229)	(0.0229)	(0.0229)
PROMOTED	0.0154	0.1577	0.1438	0.1887
	(0.1250)	(0.3373)	(0.3372)	(0.3411)
DERBY	0.2987***	0.2652***	0.2645***	0.2639***
	(0.0603)	(0.0609)	(0.0611)	(0.0611)
FIRST	0.2903***	0.2970***	0.2950***	0.2966***
	(0.0882)	(0.0878)	(0.0879)	(0.0879)
$AVGGOAL_H$	-0.0513	-0.0276	-0.0271	-0.0239
	(0.0631)	(0.0628)	(0.0629)	(0.0632)
$AVGGOAL_A$	0.0587	0.0579	0.0572	0.0516
	(0.0457)	(0.0557)	(0.0558)	(0.0572)
$AVGCONCED_H$	-0.0696	-0.0747	-0.0746	-0.0780
	(0.0613)	(0.0612)	(0.0613)	(0.0617)
$AVGCONCED_A$	-0.0807*	-0.1095*	-0.1079*	-0.1030*
	(0.0474)	(0.0558)	(0.0558)	(0.0570)
$AVGPOINTS_H$	0.0737	0.0504	0.0508	0.0509
	(0.0719)	(0.0720)	(0.0721)	(0.0721)
$AVGPOINTS_A$	-0.0177	-0.0325	-0.0318	-0.0306
	(0.0535)	(0.0640)	(0.0640)	(0.0641)
$HABIT_H$	0.0184	0.1440	0.1100	0.2281
	(0.1605)	(0.9498)	(0.9497)	(0.9610)
$HABIT_A$	0.0984***	0.0628	0.0611	0.0668
	(0.0216)	(0.0725)	(0.0725)	(0.0728)
RANK	-0.0341	-0.0218	-0.0197	-0.0256
	(0.0326)	(0.0352)	(0.0351)	(0.0356)
Home-FE	YES	YES	YES	YES
Away-FE	NO	YES	YES	YES
Time-FE	YES	YES	YES	YES
R-squared	0.9013	0.9064	0.9063	0.9065
Observations	704	704	704	704

Notes: Standard Errors in parentheses. *, **, *** denote statistical significance at the 0.10, 0.05, 0.01 level.

Table 4 Robustness Checks

	ln Attendance			
	(1)	(2)	(3)	(4)
$PROB_H$	-0.8553*	-0.7633*	-0.7550*	-0.8454*
	(0.4540)	(0.4544)	(0.4526)	(0.4833)
$PROB_H^2$	0.8376*	0.8202*	0.8183*	0.9044*
	(0.4559)	(0.4472)	(0.4464)	(0.4724)
WEEKEND	0.0420*	0.0410*	0.0415*	0.0428**
	(0.0219)	(0.0219)	(0.0219)	(0.0217)
PROMOTED	0.1794	0.1744	0.1651	0.2069
	(0.1365)	(0.3219)	(0.3231)	(0.3342)
DERBY	0.2530***	0.2602***	0.2608***	0.2643***
	(0.0562)	(0.0553)	(0.0554)	(0.0557)
FIRST	0.3003***	0.2975***	0.2973***	0.2974***
	(0.0970)	(0.0915)	(0.0919)	(0.0909)
$AVGGOAL_H$		0.0146	-0.0244	-0.0246
		(0.0383)	(0.0677)	(0.0679)
$AVGGOAL_A$		0.0283	0.0453	0.0517
		(0.0351)	(0.0555)	(0.0545)
$AVGCONCED_H$		-0.1042*	-0.0733	-0.0767
		(0.0549)	(0.0654)	(0.0653)
$AVGCONCED_A$		-0.0830	-0.1012	-0.1026
		(0.0541)	(0.0783)	(0.0783)
$HABIT_H$		0.1564	0.1266	0.2680
		(0.9562)	(0.9560)	(0.9898)
$HABIT_A$		0.0586	0.0622	0.0693
		(0.0707)	(0.0719)	(0.0726)
$AVGPOINTS_H$			0.0538	0.0525
			(0.0748)	(0.0748)
$AVGPOINTS_A$			-0.0242	-0.0308
			(0.0658)	(0.0644)
RANK				-0.0282
				(0.0354)
Home-FE	YES	YES	YES	YES
Away-FE	YES	YES	YES	YES
Time-FE	YES	YES	YES	YES
R-squared	0.9041	0.9064	0.9065	0.9066
Observations	704	704	704	704

Notes: Standard Errors in parentheses. *, **, *** denote statistical significance at the 0.10, 0.05, 0.01 level.

8 Discussion

The aim of this study is to examine the effect of the expected outcome on attendance in Swedish Allsvenskan. The empirical analysis is based on the behavioural model proposed by Coates, Humphreys, & Zhou (2014). Our results are statistically significant and show that a U-shaped relationship exists between attendance and the probability of a home team win. This section discusses to what extent the result answers the research question. First, it explains the main findings of the study. Next, it comments on the validity of the results and relates them to previous research. Finally, it discusses the implications of the results.

The empirical results provide evidence that behavioural economic concepts are relevant for analysing attendance demand in football games, in particular the concepts of reference dependence and loss-aversion. The estimated coefficients of the expected outcome variables, $PROB_H$ and $PROB^2_H$, show the average effect the variables have on game attendance, controlling for differences in potential attendance between games. Given the assumptions we make about fans, the coefficients also show the preferences of the marginal fan. The main result suggests that across all teams, fans have on average the lowest expected utility at the win probability 0.46 and the highest expected utility at the win probability 0 or 1. Thus, fans show a preference for games with little outcome uncertainty. The most likely explanation for the exhibited preferences is that fans are loss-averse, which dominates the marginal consumption utility of a home win. In other words, for a relatively certain loss, the expected consumption utility is smaller than for a more uncertain outcome, but this is compensated by a less negative expected gain-loss utility. In addition, for a relatively certain win, the expected consumption utility is greater than for a more uncertain outcome, and this is enforced by a less negative-expected gain-loss utility.

These findings might be somewhat limited in terms of internal validity. There may be a bias due to omitted factors, in particular time-variant heterogeneity. This would mean that the coefficients are not true parameters. Nevertheless, when running our robustness check and sensitivity analysis, the signs do not change. Therefore, our results are robust qualitatively, that is, the overall feature of the relationship is likely to hold.

In terms of the external validity of this study, we want to emphasize the problems of generalising the results outside the context of *Swedish Allsvenskan*. Take for example US sports; their league is set up in a completely different way than Swedish *Allsvenskan*. This applies to the administrative structure as well as the competition format

Furthermore, the main result is broadly consistent with the trends in the literature. The rejection of Rottenberg's uncertainty of outcome hypothesis is in line with most research on the highest league of European football. Regarding the general literature on football attendance, this study has limited success in replicating results. The results for DERBY, WEEKEND and FIRST is in line with previous research, but other control variables do not appear to affect attendance.

The study raises questions about the level of competition between teams in Swedish Allsvenskan. The empirical results show that very uncertain games have almost 20 % less attendance than games where one team clearly dominates. This difference implies a noticeable reduction in revenues for teams. The absolute reduction is greater for teams with higher attendance. However, the effect size is rather small compared to the differences in attendance between teams. It seems like potential attendance plays a more important role in game attendance than the level of competition between teams. In addition, we do not know how potential attendance is influenced by seasonal uncertainty or league uncertainty. For example, if one team dominates the league this increases attendance on the game level, but it is not clear how attendance is affected if this team wins the league every season. Thus, the study does not provide an unambiguous recommendation. A change in the competition format that leads to a decrease in the competitiveness of the league might not lead to an increase in attendance.

To sum up, in Swedish Allsvenskan attendance is on average the highest for games with little outcome uncertainty. This indicates that fans have an aversion for an unexpected loss, which dominates the marginal utility of a home team win. The overall feature of the relationship has strong internal validity. Due to structural differences between sports, it is questionable if the results can be generalised to contexts outside of the highest leagues of European football. This view is also supported by the literature. The findings do not imply that Swedish Allsvenskan can gain more attendance by decreasing competition. Instead, further research is needed on seasonal uncertainty and league-level uncertainty.

9 Conclusion

In this study, we analyse panel data on football attendance in Swedish Allsvenskan, and this is guided by a behavioural model of the decision to attend sports events. The goal is to answer the research question *'what is the effect of the expected outcome on attendance for football games?´*.

The main feature of the behavioural model is that overall utility consists of two parts. One part is an outcome-based consumption utility. The other part is a gain-loss utility which is determined relative to a reference point, and the reference point is based on rational expectations about the game outcome. Depending on what assumptions are made about fan preferences, the behavioural model gives several predictions about the relationship between expected outcome and game attendance. If fan preferences are characterised by loss aversion, then the relationship between the home team's probability of winning and game attendance is convex. On the other hand, if fan preferences give a concave relationship, then the uncertainty of outcome hypothesis holds. The empirical strategy estimates the relationship between the home team win probability and game attendance, and the estimated parameters indicate if fans show preferences in line with the theoretical predictions. To alleviate bias, we use a fixed effects estimator.

The results suggest that the uncertainty of outcome hypothesis does not hold in our application, instead the opposite seems to be true. So, in Swedish Allsvenskan fans seem to prefer certain outcomes due to loss aversion. We likely have some biases that affect the estimated coefficient parameters. However, we argue that our results are qualitatively robust, that is, the overall feature of the U-shaped relationship between attendance and home win probability is likely to hold. These results provide further support for the relevance of prospect theory. Moreover, our findings imply that quality differences between teams in Allsvenskan might be advantageous for attendance. However, further research should investigate if this conclusion holds when the impact of long-term uncertainty on attendance is also taken into consideration.

References

Allsvenskan. (2021). Om oss, Available online: https://www.allsvenskan.se/om-oss [Acessed 1 May 2021]

Baranzini, A., Ramirez, J., & Weber, S. (2008). The Demand for Football in Switzerland: An Empirical Estimation. Available online: https://ssrn.com/abstract=1087243 [Accessed 16 May 2021]

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. 2, pp. 216–235.

Besters, L. M., van Ours, J. C., & van Tuijl, M. A. (2019). How outcome uncertainty, loss aversion and team quality affect stadium attendance in Dutch professional football, *Journal of Economic Psychology*, Vol. 72, pp. 117-127.

Borland, J., & MacDonald, R. (2003). Demand for sport. *Oxford review of economic policy*, *19*(4), 478-502.

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, pp. 326-338.

Card, D., & Dahl, G. B. (2011). Family violence and football: the effect of unexpected emotional cues on violent behavior, *The quarterly journal of economics*, vol. 126, no. 1, pp. 103–143.

Coates, D., Humphreys, B. R., & Zhou, Li. (2012). Outcome Uncertainty, Reference-Dependent Preferences and Live Game Attendance, Working papers 2012-7, University of Alberta, Department of Economics. Available online: https://ideas.repec.org/p/ris/albaec/2012_007.html [Accessed 11 May 2021]

Coates, D., Humphreys, B. R., & Zhou, L. (2014). Reference-dependent preferences, loss aversion, and live game attendance, *Economic Inquiry*, vol. 52, no.3, pp. 959-973.

Cox, A. (2018). Spectator Demand, Uncertainty of Results, and Public Interest: Evidence From the English Premier League, *Journal of Sports Economics*, vol. 19, no. 1, pp. 3–30.

Di Domizio, M., & Caruso, R. (2015). Hooliganism and Demand for Football in Italy: Attendance and Counterviolence Policy Evaluation, *German Economic Review*, vol. 16, no. 2, pp. 123-137._

Eckard, E. W. (2017) The Uncertainty-of-Outcome Hypothesis and the Industrial Organization of Sports Leagues: Evidence From U.S. College Football, *Journal of Sports Economics*, vol. 18, no. 3, pp.298-317.

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), 51*(2), 229-241.

Humphreys, B. R., & Zhou, L. (2015) The Louis-Schmelling paradox and the league standing effect reconsidered, *Journal of Sports Economics*, vol. 16, no. 8, pp. 835-852

Jespersen, N. D., & Pedersen, L. B. (2018). Rejecting the Uncertainty of Outcome Hypothesis on Attendance Demand in all Four Major European Football Leagues, *Forum for Idræt*, vol. 33, no. 1, pp. 100–116.

Kahneman, D., & Tversky, A. (1979). Prospect theory: An analysis of decision under risk, *Econometrica*, vol. 47, no. 2, pp. 263–292.

Koszegi, B., & Rabin, M. (2006). A model of reference-dependent preferences, *The Quarterly Journal of Economics*, vo. 121, no.4, pp. 1133.

Kuypers, T. (2000). Information and efficiency: an empirical study of a fixed odds betting market, *Applied Economics*, vol. 32, no. 11, pp. 1353-1363.

Lahvicka, J, (2013) Does Match Uncertainty Increase Attendance? A Non-Regression Approach, Available online: https://ssrn.com/abstract=2297320 [Accessed 16 May 2021]

Martins, A. M., & Cró, S. (2018). The Demand for Football in Portugal: New Insights on Outcome Uncertainty, *Journal of Sports Economics*, vol. 19, no. 4, pp. 473–497.

Mills, B., & Fort, R. (2014). League-level attendance and outcome uncertainty in U.S. Pro Sports Leagues, *Economic Inquiry*, vol. 52, pp. 205–218.

Neale, W. C., (1964). The Peculiar Economics of Professional Sports, *Quarterly Journal of Economics*, Vol. 78, No. 1, pp. 1-14.

Paul, R. J., Wachsman, Y., Weinbach, A. P. (2010). The role of uncertainty of outcome and scoring in the determination of fan satisfaction in the NFL, *Journal of Sports Economics*, vol. 12, pp. 213–221.

Pawlowski, T., & Anders, C. (2012). Stadium attendance in German professional football – the (un)importance of uncertainty of outcome reconsidered, *Applied Economics Letters*, vol. 19, no .16, pp. 1553-1556

Peel, D. A., & Thomas, D. A. (1992). The demand for football: Some evidence on outcome uncertainty. *Empirical Economics*, *17*(2), 323-331.

Reilly, B. (2015). The Demand for League of Ireland Football, *The Economic and social review*, vol. 46, pp. 485-509.

Rottenberg, S. (1956). The baseball player's labour market, *Journal of Political Economy*, vol. 64, pp. 242–58.

Roy, P. (2004). Die Zuschauernachfrage im professionellen Teamsport: Eine ökonomische Untersuchung am Beispiel der deutschen Fußball-Bundesliga. Aachen: Shaker-Verl.

Sacheti, A., Gregory-Smith, I., Paton, D. (2014). Uncertainty of outcome or strengths of teams: An economic analysis of attendance demand for international cricket, *Applied Economics*, vol. 46, pp. 2034–2046.

Schreyer, D., Schmidt, S. L., & Torgler, B. (2018). Game outcome uncertainty in the English Premier League: Do German fans care?. *Journal of Sports Economics*, *19*(5), 625-644.

Serrano, R., García-Bernal, J., Fernández-Olmos, M., & Espitia-Escuer, M. A. (2015) Expected quality in European football attendance: market value and uncertainty reconsidered, *Applied Economics Letters*, vol. 22 no. 13, 1051-1054.

Sport & Affärer. (2021). Klubbarnas varumärke. Available online: https://www.sportaffarer.se/?sida=49 [Accessed 01 May 2021]

Svenska Fotbollsförbundet. (2021). Tävlingsbestämmelser 2021, Available online: https://www.svenskfotboll.se [Accessed 29 April 2021]

Svenska Fotbollsförbundet. (2021). Allsvenskan herrar 2017, Available online: https://www.svenskfotboll.se/serier-cuper/tabell-och-resultat/allsvenskan-herrar/67049/ [Accessed 29 April 2021]

Svenska Fotbollsförbundet. (2021). Allsvenskan herrar 2018, Available online: https://www.svenskfotboll.se/serier-cuper/tabell-och-resultat/allsvenskan-herrar/72152/ [Accessed 29 April 2021]

Svenska Fotbollsförbundet. (2021). Allsvenskan herrar 2019, Available online: https://www.svenskfotboll.se/serier-cuper/tabell-och-resultat/allsvenskan-herrar/77486/ [Accessed 29 April 2021]

SVT. (2013). Fakta: 51-procentsregeln, Available online: https://www.svt.se/sport/artikel/fakta-51-procentsregeln [Accessed 29 April 2021]

Szymanski, S. (2003). The economic design of sporting contests, *Journal of Economic Literature*, vol. 41 no. 4, pp. 1137-1187.

Verbeek, M. (2008). A guide to modern econometrics. Chichester, England: John Wiley & Sons.

Wooldridge, J. M. (2016). Introductory econometrics: A modern approach. Boston, Ma.: Cengage learning.

Appendix A

From the behavioural model we derive a regression model for the relationship between expected outcome and observed attendance. The derivation exactly follows Coates, Humphreys, and Zhou (2012). First, we assume that the reservation utility follows a continuous uniform distribution, with maximum $\bar{v}$ and minimum $\underline{v}$. So, at a game where home team *i* plays against away team *j* at time *t* the attendance is

$$Attendance_{ijt} = PotentialAttendance_{ijt} \cdot \frac{E[U]_{ijt} - \underline{v}}{\overline{v} - \underline{v}}$$

Next, we take the natural logarithm on both sides of the equation and separate the right-hand side into two terms. We also use the approximation ln(x + 1) ≈ x for changing the last term:

$$\ln \frac{E[U] - \underline{v}}{\overline{v} - \underline{v}} = \ln \frac{E[U] - \underline{v} + \overline{v} - \overline{v}}{\overline{v} - \underline{v}} = \ln \frac{E[U] - \overline{v}}{\overline{v} - \underline{v}} + 1 \approx \frac{E[U] - \overline{v}}{\overline{v} - \underline{v}}$$

So, the equation is

$$\ln Attendance_{ijt} = \ln PotentialAttendance_{ijt} + \frac{E[U]_{ijt} - \overline{v}}{\overline{v} - \underline{v}}$$

Then, we plug in the expression for $E[U]_{ijt}$ from equation 2:

$$\begin{aligned}
&\ln Attendance_{ijt} \\
&= \ln PotentialAttendance_{ijt} + \frac{(\beta - \alpha)p_{ijt}^2 + [(U^W + U^L) - (\beta - \alpha)]p_{ijt} + U^L - \overline{v}}{\overline{v} - \underline{v}} \\
&= \ln PotentialAttendance_{ijt} + \frac{(\beta - \alpha)p_{ijt}^2}{\overline{v} - \underline{v}} + \frac{[(U^W + U^L) - (\beta - \alpha)]p_{ijt}}{\overline{v} - \underline{v}} + \frac{U^L - \overline{v}}{\overline{v} - \underline{v}}
\end{aligned}$$

Lastly, we assume that the functional form for potential attendance is

$$PotentialAttendance_{ijt} = e^{X_{ijt}\mu + \varepsilon_{ijt}}$$

and define $\gamma \equiv \frac{(\beta-\alpha)}{\overline{v}-\underline{v}}$, $\theta \equiv \frac{[(U^W+U^L)-(\beta-\alpha)]}{\overline{v}-\underline{v}}$, and $\lambda \equiv \frac{U^L-\overline{v}}{\overline{v}-\underline{v}}$. Thus, we have the final expression of the regression model:

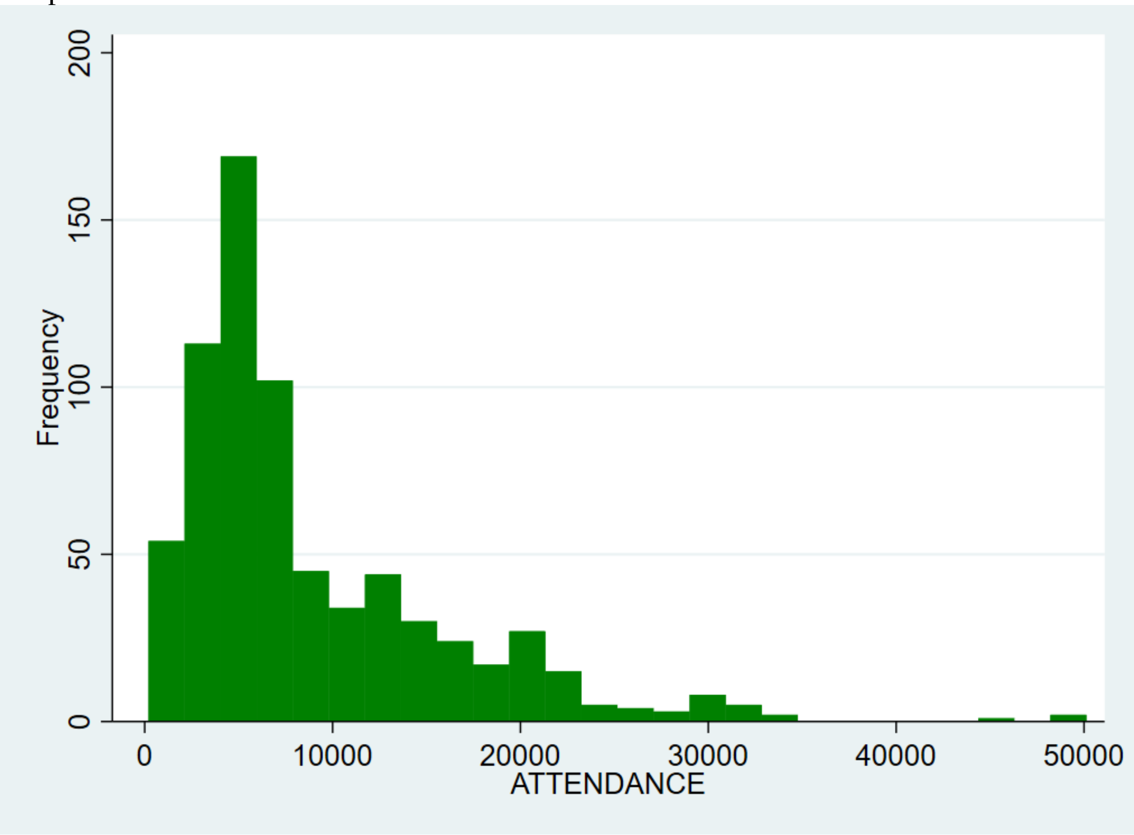
$$\ln Attendance_{ijt} = \lambda + \theta p_{ijt} + \gamma p_{ijt}^2 + X_{ijt}\mu + \varepsilon_{ijt}$$

Appendix B

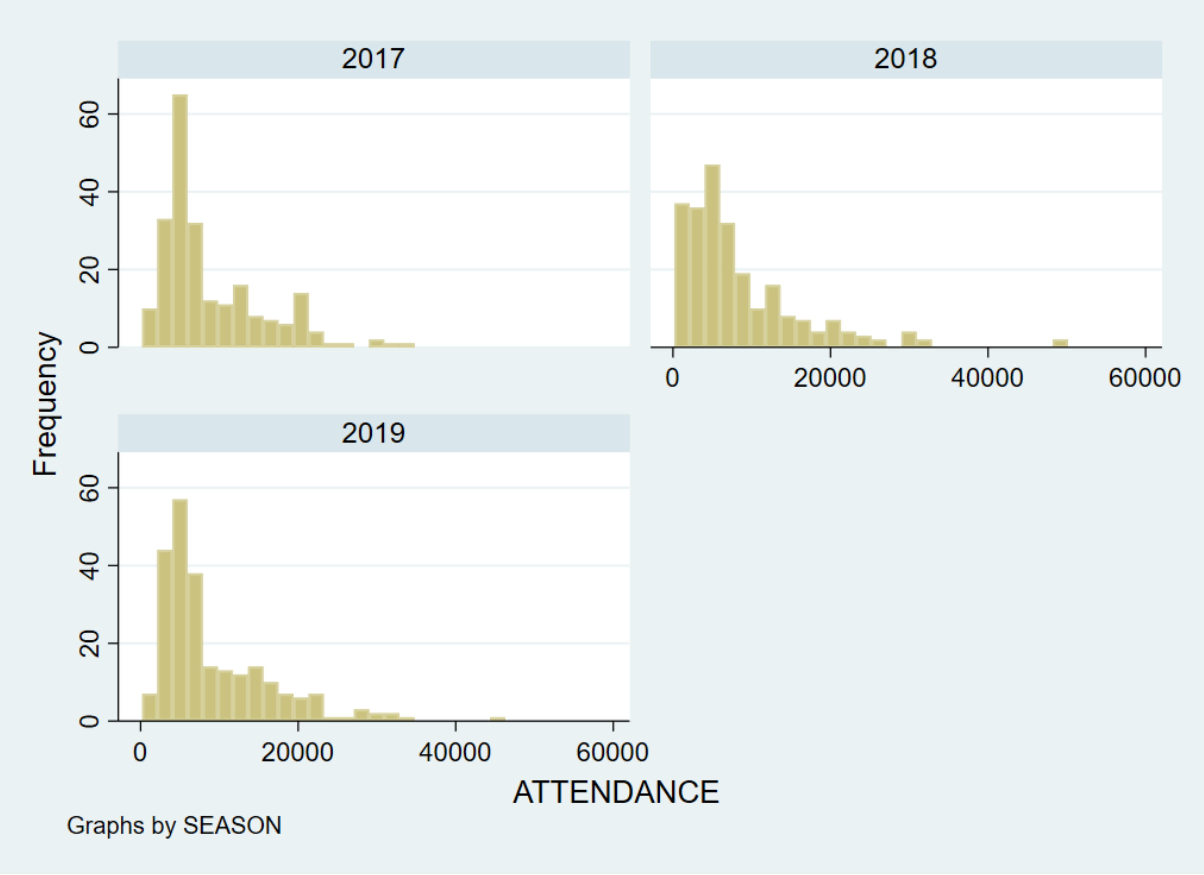
Table B1: Teams and Seasons

Team		Season	
	2017	2018	2019
AFC ESKILSTUNA	X		X
AIK	X	X	X
BK HACKEN	X	X	X
DALKURD FF		X	
DJURGARDENS IF	X	X	X
FALKENBERGS FF			X
GIF SUNDSVALL	X	X	X
HALMSTADS BK	X		
HAMMARBY IF	X	X	X
HELSINGBORGS IF			X
IF BROMMAPOJKARNA		X	
IF ELFSBORG	X	X	X
IFK GOTEBORG	X	X	X
IFK NORRKOPING FK	X	X	X
IK SIRIUS	X	X	X
JONKOPINGS SODRA IF	X		
KALMAR FF	X	X	X
MALMO FF	X	X	X
OREBRO SK	X	X	X
OSTERSUNDS FK	X	X	X
TRELLEBORGS FF		X	

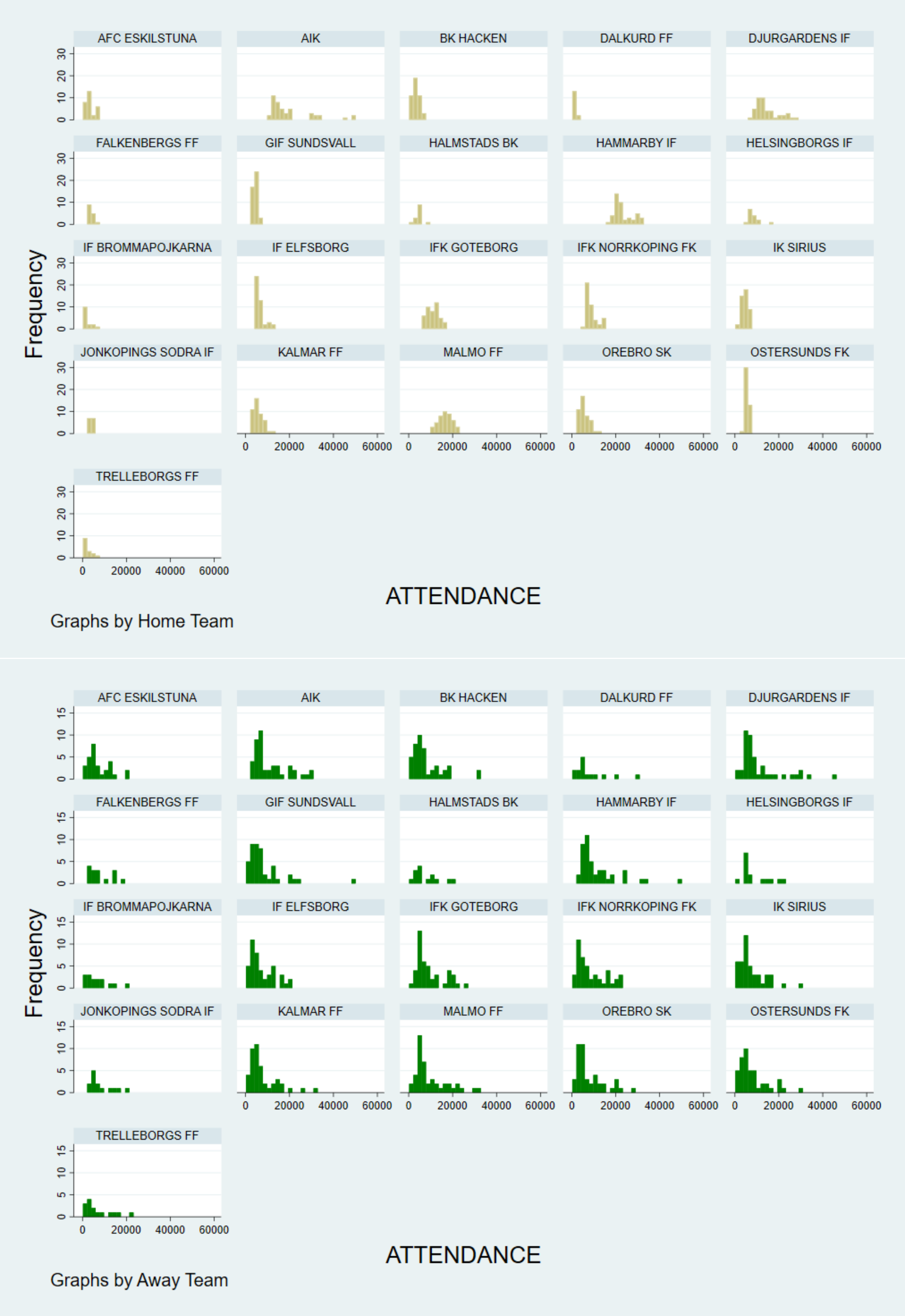
Graph B1: The Overall Distribution of Attendance



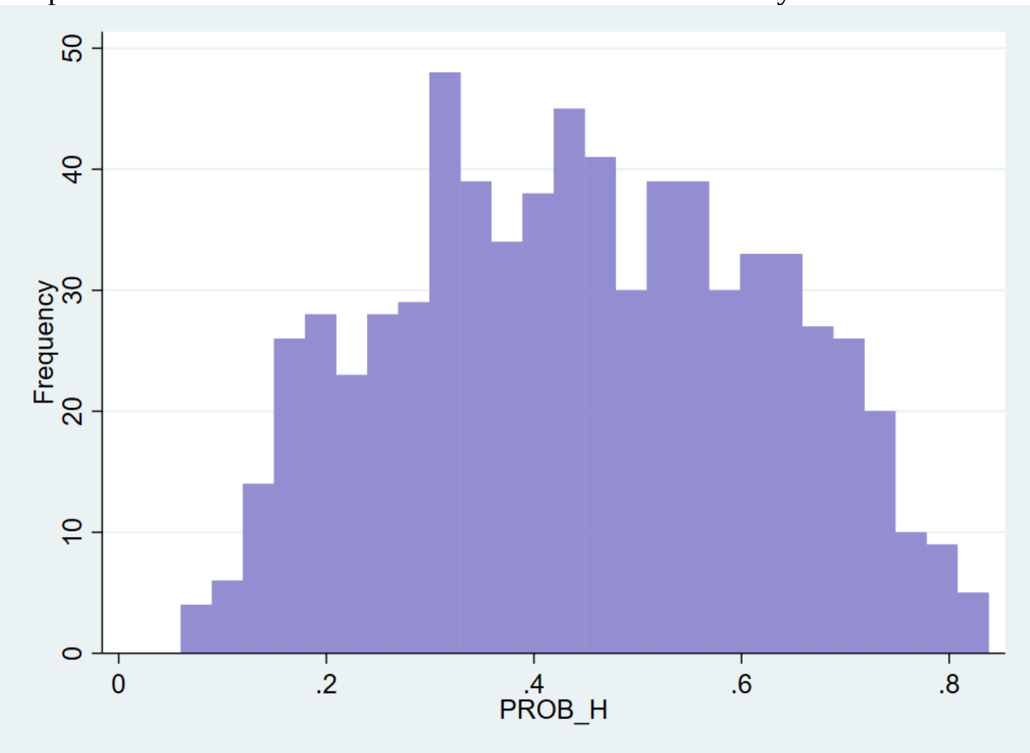
Graph B2: The Distribution of Attendance for each Season



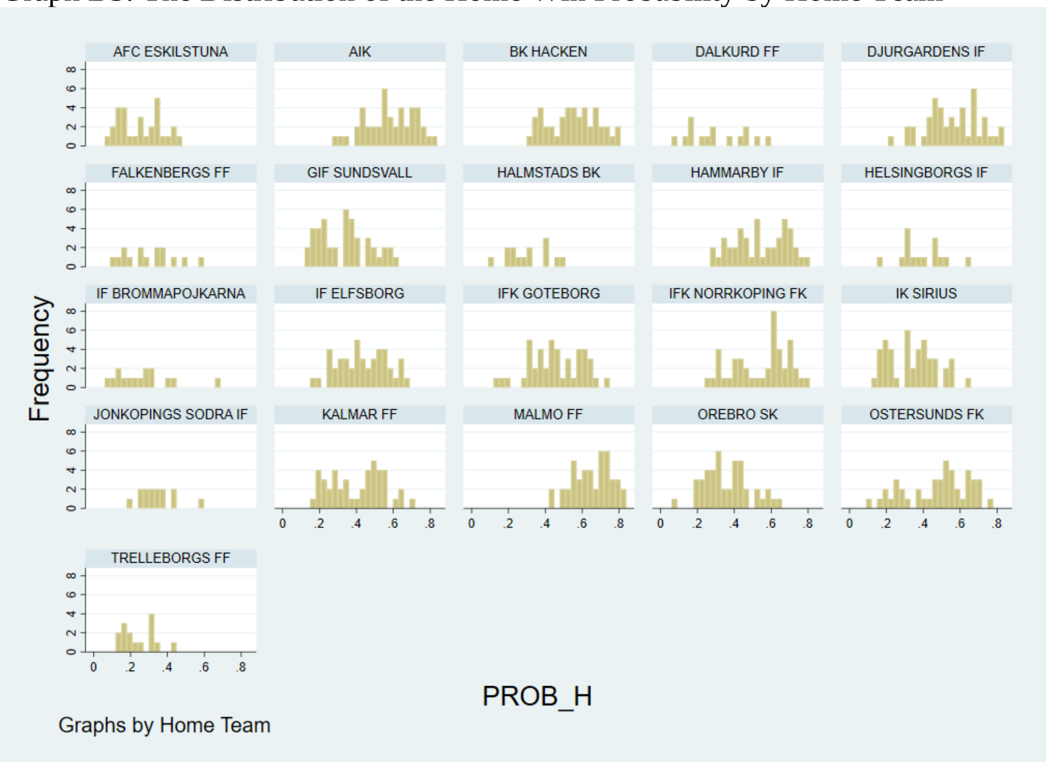
Graph B3: The Distribution of Attendance by Team for Home Games and Away Games.



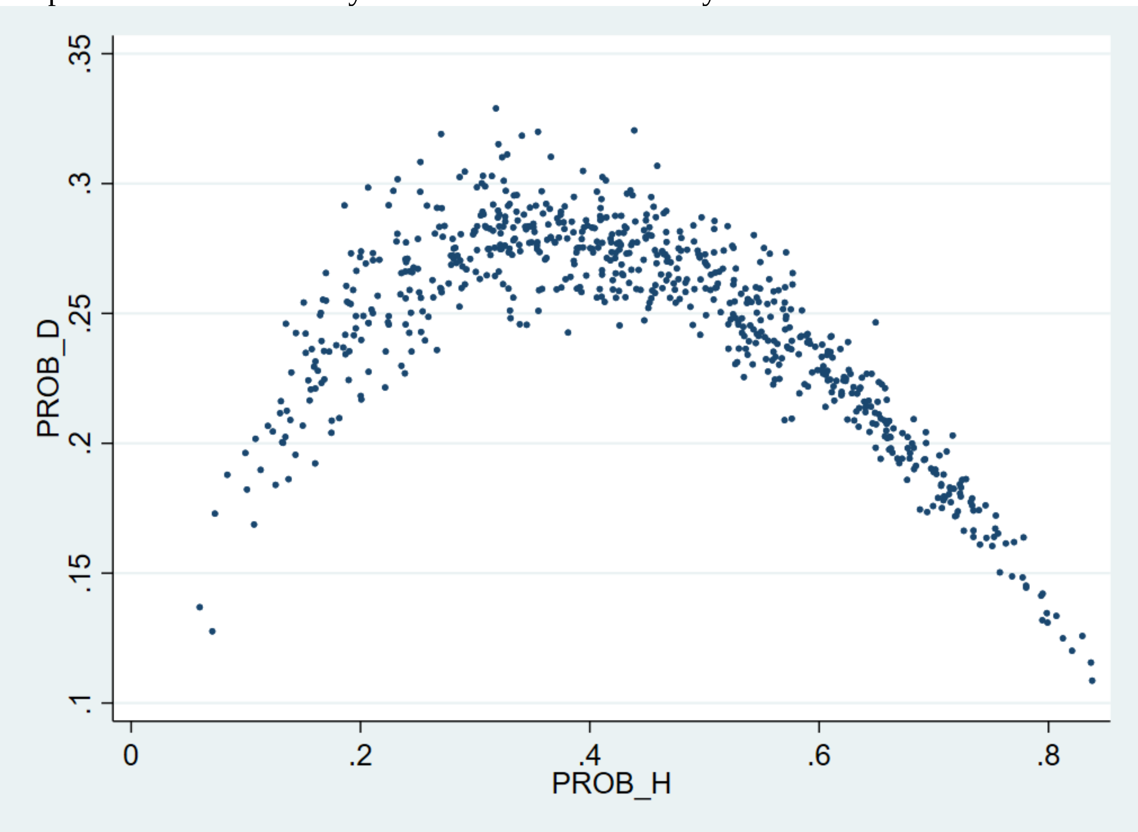
Graph B4: The Overall Distribution of Home Win Probability



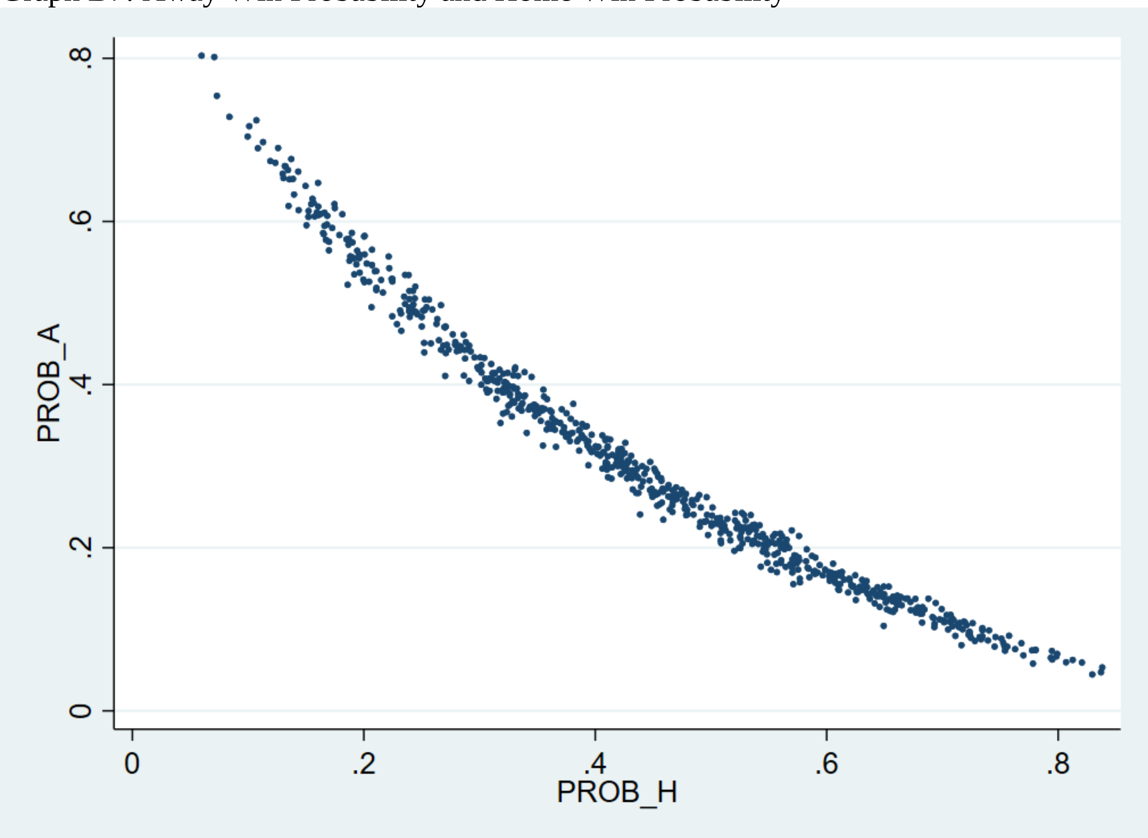
Graph B5: The Distribution of the Home Win Probability by Home Team



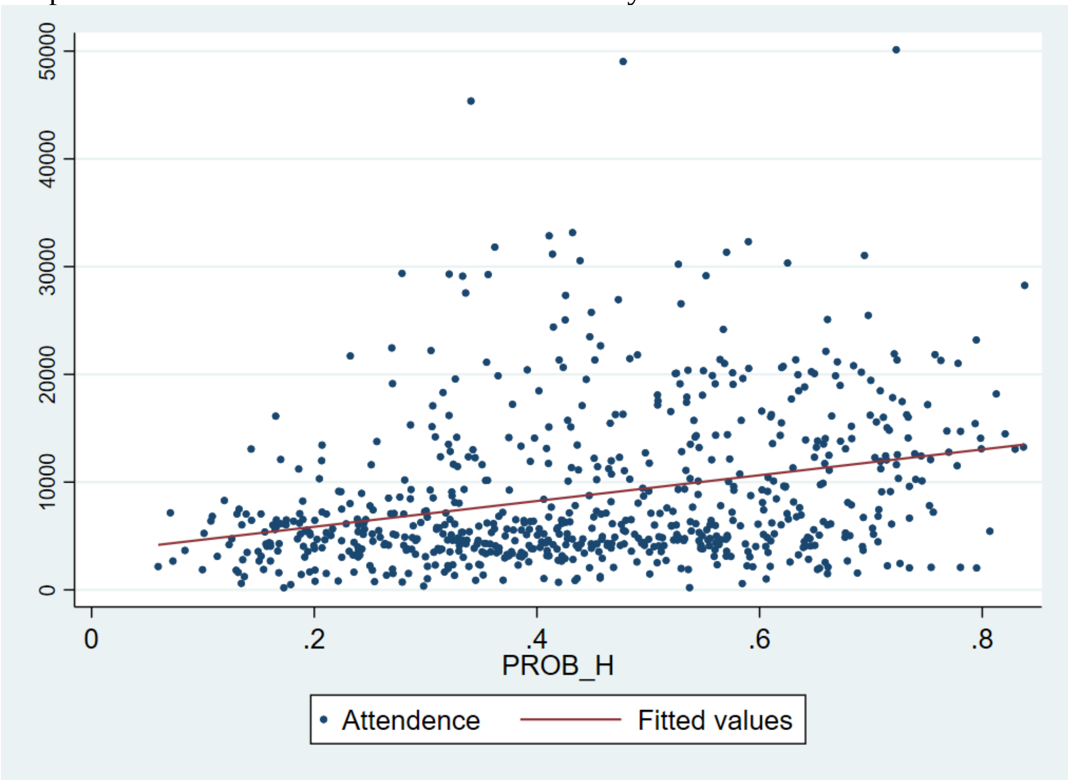
Graph B6: Draw Probability and Home Win Probability



Graph B7: Away Win Probability and Home Win Probability



Graph B8: Attendance and Home Win Probability



Graph B9: Attendance and Home Win Probability by Home Team