

The impact of the soccer schedule on TV viewership and stadium attendance: Evidence from the Belgian Pro League

Chang Wang

Faculty of Economics and Business

KU Leuven, Belgium

Dries Goossens

Faculty of Economics and Business Administration

Ghent University, Belgium

Martina Vandebroek

Faculty of Economics and Business and Leuven Statistics Research Centre

KU Leuven, Belgium

September 15, 2015

Abstract

In the past decade, television broadcasters have been investing a huge amount of money for the Belgian Pro League broadcasting rights. These companies pursue an audience rating maximization, which depends heavily on the schedule of the league matches. At the same time, clubs try to maximize their home attendance and find themselves affected by the schedule as well. Our paper aims to capture the Belgian soccer fans' preferences with respect to scheduling options, both for watching matches on TV and in the stadium. We carried out a discrete choice experiment using an online survey questionnaire distributed on a national scale. The choice sets are based on three match characteristics: month, kickoff time, and quality of the opponent. The first part of this survey concerns television broadcasting aspects. The second part includes questions about stadium attendance. The choice data is first analyzed with a conditional logit model which assumes homogenous preferences. Then a mixed logit model is fit to model the heterogeneity among the fans. The estimates are used to calculate the expected utility of watching a Belgian Pro League match for every possible setting, either on TV or in the stadium. These predictions are validated in terms of the real audience rating and home attendance data. Our results can be used to improve the scheduling process of the Belgian Pro League in order to persuade more fans to watch the matches on TV or in a stadium.

keywords: Audience ratings; Belgian soccer; Conditional logit model; Discrete choice experiment; Mixed logit model; Schedule; Stadium attendance

1 Introduction

In the past decade, television broadcasters have been investing a huge amount of money for the broadcasting rights of European professional soccer (in some countries, referred to as “football”) competitions. The English Premier League has sold their next three seasons’ television broadcasting rights to *British Telecom* and *Sky* for £3 billion ¹, a new record in the professional soccer world. A rapid increase in television broadcasting prices can also be observed in many other European countries, including Belgium. In 2005, *Belgacom* paid €36 million for the rights for live broadcasting of all the Belgian Pro League matches for the next three seasons. In 2014, the broadcasting rights for one season were sold for €70 million, non-exclusively allocated to three companies. Given this huge cost, television broadcasters definitely seek to maximize their revenue by attracting as many viewers as possible. At the same time, clubs try to maximize their home attendance. Indeed, the matchday revenue, which is largely derived from gate receipts (including season tickets and memberships), is one of the main sources of a club’s total revenue. According to Deloitte Football Money League 2014 ², the proportion of the matchday revenue among the top 20 European clubs in this league varies from 10% (AC Milan) to 38% (Arsenal F.C.). Furthermore, this source of revenue is becoming increasingly important, given the UEFA Financial Fair Play regulations, which forces clubs to live within their means (Peeters & Szymanski 2014).

In this text, we study the impact of the schedule on TV viewership and stadium attendance for the Belgian Pro League. We want to figure out what scheduling options attract Belgian fans, and capture their preferences for watching a match on TV and attending it in the stadium. We also analyze whether these preferences are similar, or if broadcasters and clubs indeed have competing interests with respect to the schedule. The results of our study are valuable inputs for the scheduling process, and should lead to fixtures that persuade more fans to watch the matches.

The number of studies on TV viewership for soccer is rather limited. Alavy et al. (2010) focus on the relation between viewer ratings and match outcome uncertainty. Forrest et al. (2005) also mainly focus on outcome uncertainty, and use their model to support the broadcasters’ choice of which matches to televise in the English Premier League. Buraimo (2008) models the audience ratings for the second tier of English league soccer, taking into account team form, player talent, outcome uncertainty, and atmosphere. Further, Feddersen and Rott (2011) identify determinants of demand for televised matches of the German national team. Demand studies of professional soccer have traditionally focused on stadium attendance (see Borland & McDonald (2003) for an overview). Various demand models have been used to analyze the factors that determine stadium attendance, taking into account some league-specific dummy variables (e.g., Garcia & Rodriguez 2002; Lynch 2005; Madalozzo &

¹<http://www.dailymail.co.uk/sport/football/article-2158825>

²<http://www.deloitte.com/assets/Dcom-UnitedKingdom/Local%20Assets/Documents>

Villar 2009). Compared to television broadcasting, more factors can influence a soccer fan's decision whether or not to travel to the stadium. Indeed, fans usually also take into account the weather, distance from the stadium, availability of tickets, availability of car parking, safety, service, and catering, when making a decision (Carvalho et al. 2013). A few studies have focused on the impact of stadium attendance on the size of the television audience, and vice versa. For English league soccer, Buraimo (2008) found that stadium attendance has a positive impact on television audience, but the reverse impact is negative. He also argued that stadium fans tend to be loyal supporters and a large proportion of them have bought season tickets, while television viewers tend to be less devoted to their team. Other researchers have found that television broadcasting has either zero effect or a negative effect on the demand for tickets in European soccer leagues (Baimbridge et al. 1996; Allan & Roy 2008). In each of these studies, the schedule has been considered as given. As far as we are aware, there is no literature focusing on how TV viewership and stadium attendance are affected by the schedule in domestic league soccer.

Essentially, a schedule determines which opponents will face each other in which part of the season, and at what kickoff time. Hence, we consider three match characteristics: the month, the kickoff time, and the strength of the opponent. One obvious way to understand the impact of the schedule on TV viewership (or stadium attendance) would be to study historical viewing (or attendance) figures. Even if we had these data available for sufficiently many past seasons, it would present an incomplete picture, since some match and kickoff time combinations have almost never been used in the Belgian Pro League. For instance, Saturday 20:00 has hardly ever been used for top matches and Sunday 14:30 has rarely been used for a relegation clash. In this study, we used an online survey to gather the necessary information on the Belgian soccer fans' preferences in watching the league matches. The first part of the survey contains questions concerning television broadcasting issues. The second part includes questions about stadium attendance. In each part, some background questions are followed by a set of choice questions, which are based on a semi-Bayesian conditional logit D-optimal design (Sandor & Wedel 2001; Kessels et al. 2006). The collected choice data is first analyzed with a conditional logit model, which assumes homogenous preferences. Then a mixed logit model is fit to the data to model the heterogeneity among the respondents. The estimates are used for making predictions, which are compared with the real audience rating and home attendance data from the 2013–14 season of the Belgian Pro League. Our method can easily be applied to soccer competitions in other countries.

The paper is organized as follows. Section 2 gives a brief introduction to the Belgian Pro League and current scheduling practices. Section 3 provides details about the construction of the choice sets of the questionnaire and introduces the two statistical models that were estimated. In Section 4 we discuss the results, including parameter estimation, predictions, and validation by means of the real TV ratings and stadium attendance data. Section 5 discusses the applicability of our results to the scheduling process. We make some concluding remarks in Section 6.

2 The Belgian Pro League

The Belgian Pro League is the top league competition for soccer clubs in Belgium. Compared with most other domestic leagues in Europe, Belgium uses a rather complicated competition format, consisting of a regular and a play-off stage (Goossens & Spieksma 2012). The regular stage has sixteen teams competing in a double round robin tournament. It usually starts in late July and ends next March (including a winter break in January). Depending on the ranking a team obtains in the regular stage, it is assigned to one of three play-off competitions, where it plays for the league title, qualification for European club competitions, or simply to avoid being relegated. Depending on its performance, a team in the Belgian Pro League plays between 33 and 42 matches per season in the domestic league. Typically, matches are scheduled on weekends, although occasionally midweek rounds are used as well.

Since 2006, the schedule for the Belgian Pro League has been computed using a phased approach based on mixed integer programming (Goossens & Spieksma 2009). This method allows taking into account various types of wishes and constraints, each with their own weight according to their relative importance. These wishes originate from the stakeholders, such as the clubs, broadcasters, and the police. For instance, the mayor of a city may forbid a home match with a high risk of hooliganism to be scheduled on a particular date if the local police force is not able to guarantee safety at that time (e.g., because of other events). Wishes specified by clubs generally aim to increase their stadium attendance and/or improve their chances for sporting success. For all teams, it is highly important to avoid successions of home matches. This is supported by Forrest and Simmons (2006), who found that scheduling home matches close to one another has an adverse impact on attendance in the English Premier League. Several teams like to play the first match at home, as this increases their attendance given the high media coverage for the opening round, and at the same time increases their chances of starting the season with a win. Some teams prefer to avoid top matches in the holiday period, or playing a home match simultaneously with some other local event. Not all teams seem to agree on how to maximize attendance: some teams wish to play home matches against top teams in the first half of the season, whereas others prefer to host top matches as late as possible in the season.

Before 2005, home teams had the right to decide on the actual time at which their matches were played. Almost all matches were played either on Saturday at 20:00, or on Sunday at 15:00. As live broadcasting of matches was started in 2005, the broadcaster requested a larger variety in kickoff times to attract more viewers. As a result, four time slots were typically used for weekend matches: Friday 20:30, Saturday 20:00, and Sunday 18:00 and 20:30. In the subsequent years, the competition between several Belgian broadcasting companies to bid for the rights intensified, and the price went up. As a result, more time slots were used, and the TV broadcaster had an increasingly important say in what match to play on what round, and at which kickoff time. Currently, *Belgacom*, *Telenet* and *VOO* hold the TV rights non-exclusively. Live matches from the Belgian Pro League are being broadcasted in the Netherlands, Hungary, Romania, Serbia, and Turkey.

3 Methodology

This study aims to capture Belgian soccer fans’ preferences in watching the Belgian Pro League using an online survey based on a discrete choice experiment. In the literature, discrete choice experiments are frequently used to study consumer choice behavior. One reason for their popularity is that they enable researchers to model realistic choices, to simulate realistic market decisions, and to predict the demand for different products (Carson et al. 1994). In a typical discrete choice experiment, respondents are presented with a series of choice sets, each consisting of several alternatives which have combinations of attribute levels. Respondents are asked to choose their preferred alternative in each choice set. One convenient measure of the relative preference of an alternative is its utility, of which the deterministic part is assumed to be a linear function of the attribute levels. Respondents are assumed to choose the alternative with the highest utility.

3.1 Discrete choice experiment: The specifications

The Belgian Pro League survey we report on consists of two separate parts. The first part contains questions about television broadcasting and the second part concerns aspects of stadium attendance. In each part, several questions are asked to get some background information. Among other things, respondents are asked to indicate their favorite club in the Belgian Pro League 2013–14 season.

The main body of the television broadcasting survey consists of a discrete choice experiments with 20 choice sets. These choice sets are shown one at a time and consist of three (hypothetical) Belgian Pro League matches of their favorite team broadcasted on TV and the respondents are asked to choose which of these matches they would prefer to watch. Each of these alternative matches are described by three attributes or characteristics. In Table 1 we show a possible choice set. In appendix A the complete choice sets used are provided.

Table 1: Example of a choice set for retrieving fan preferences for watching matches on TV.

	Option 1	Option 2	Option 3
Month	March	January	March
Kickoff time	Wednesday 20:30	Saturday 18:00	Saturday 20:00
Opponent	R.S.C. Anderlecht	OH Leuven	K.A.A. Gent

The first attribute is the “month” in which the match is scheduled. We have chosen August, October, January, and March as representatives of the four quartiles of a regular season and the corresponding buildup of suspense. Moreover, they can be associated with the weather conditions in summer, fall, winter, and spring, respectively. The second attribute is the “kickoff time” of the match, which is a combination of day and hour with 7 levels: Wednesday 20:30, Friday 20:30, Saturday 18:00, Saturday 20:00, Sunday 14:30, Sunday 18:00 and Sunday 20:30. These are the 7 most common kickoff times in the Belgian Pro League. The third attribute is the “opponent” of the respondent’s indicated favorite team. Based on the overall performance in the most recent six Belgian Pro League seasons, the 16 clubs playing in the Pro League 2013–14 season can

be categorized into three competitive levels: the top level (4 clubs), the middle level (5 clubs), and the lower level (7 clubs). We have chosen one club from each category as a representative opponent to be displayed to respondents: R.S.C. Anderlecht for the top level clubs, K.A.A. Gent for the middle level clubs and OH Leuven for the lower level clubs. When the respondent’s favorite team is one of these three representatives, we replaced that representative in the choice sets with another team from the same competitive level. In each choice question, the respondent is asked to choose the most interesting match for watching on TV.³ Given this specification, there are in all 84 candidate profiles or matches. How the 20 choice sets were chosen from the 95284 possible choice sets is explained in section 3.3.

Note that in this survey all the matches are assumed to be played between the respondent’s favorite team and one of these three representative teams. The primary reason for doing this is that a respondent is likely to lose interest quickly when choosing between matches where their favorite team is not involved at all. Another reason is that a large proportion of our targeted respondents are die-hard fans, who will always prefer the match concerning their favorite team, irrespective of the other attribute settings.

In the second part of the survey we collected information on the preferences related to stadium attendance. Similar to the structure of the television broadcasting survey, this part includes 16 choice sets but these now contain only two alternative matches each. The number of alternatives was reduced in this part because we included an extra attribute and wanted to limit the difficulty of the choice task. In Table 2 we show a possible choice set. In appendix B the choice sets used and the instructions given to the respondents are shown in more detail.

Table 2: Example of a choice set for retrieving fan preferences for stadium attendance.

	Option 1	Option 2
Month	August	October
Kickoff time	Sunday 20:30	Friday 20:30
Opponent	R.S.C. Anderlecht	K.A.A. Gent
Ticket type	Roofed standing place	Non-roofed seat place

The first three attributes taken into account in this survey are exactly in line with those of the television broadcasting survey. The fourth attribute refers to the “ticket type,” which has four levels: roofed/non-roofed, standing/seats⁴. The four attribute levels jointly constitute 336 candidate profiles and 56280 possible choice sets from which to select 16 sets. In this survey, we assumed that all the matches involved are scheduled to be played on the home field of the respondent’s favorite team. The respondent is asked to choose the match which of the 2 home matches shown he prefers to watch in the stadium. In the remainder of this paper however, we only make use of the first three attributes for estimating the parameters and making predictions and this for two reasons. First, the ticket type is irrelevant to scheduling. Second, there is a wide diversity among the 16 clubs’ home stadiums in infrastructure which cannot be taken into account.

³A short-hand notation for this choice experiment problem is $4 \times 7 \times 3/3/20$.

⁴A short-hand notation for this choice experiment is $4 \times 7 \times 3 \times 4/2/16$.

To select the choice sets that allow to estimate the statistical model efficiently, we use a semi-Bayesian D-optimal design as explained in section 3.3. However, such an optimal design can only be constructed with prior knowledge of the parameters (Atkinson & Donev 1992; Sandor & Wedel 2001, 2005). To obtain prior information and to check the feasibility of the choice tasks, we conducted first a pilot experiment based on a nearly orthogonal design. This version of the questionnaire was distributed online to all the personnel of the Faculty of Economics and Business, KU Leuven, and 60 completed survey responses were collected. We fitted the conditional logit model to these choices and the estimated parameter values were used as the prior means to compute the semi-Bayesian conditional logit design. In the following section we give a brief overview of the methodology used.

3.2 The conditional logit model

The model most often used to analyze the data from discrete choice experiments is the conditional logit model or multinomial logit model (McFadden 1974). This model assumes that all the respondents attach the same weight to an attribute level. This means that all fans are assumed to give the same weight to all attribute levels and that individual differences are not modelled but are captured in the error term. In both experiments, all attributes are effects-coded as shown in Table 3. To represent k levels, we need $k - 1$ variables.

Table 3: Effects-coded attributes in both experiments.

	month				kickoff time						opponent		
	M1	M2	M3		K1	K2	K3	K4	K5	K6	Low	O1	O2
Aug	1	0	0	Wed 20:30	1	0	0	0	0	0	Low	1	0
Oct	0	1	0	Fri 20:30	0	1	0	0	0	0	Middle	0	1
Jan	0	0	1	Sat 18:00	0	0	1	0	0	0	Top	-1	-1
Mar	-1	-1	-1	Sat 20:30	0	0	0	1	0	0			
				Sun 14:30	0	0	0	0	1	0			
				Sun 18:00	0	0	0	0	0	1			
				sun 20:30	-1	-1	-1	-1	-1	-1			

Notice that for each attribute (e.g., Month), there is a reference level (March) for which the parameter cannot be estimated directly but has to be computed as minus the sum of the parameters of the other dummy variables. So we need to estimate three parameters for “month,” six parameters for “kickoff time” and two parameters for “opponent level,” 11 parameters in total, in both the television broadcasting experiment and the stadium attendance experiment.

The random utility of respondent n choosing alternative k in choice set s is modeled as follows

$$U_{ksn} = \mathbf{x}'_{ksn} \boldsymbol{\beta} + \epsilon_{ksn}, \tag{1}$$

where \mathbf{x}_{ksn} is a q -dimensional vector describing the attribute levels of alternative k in choice set s for respondent n , and $\boldsymbol{\beta}$ is a q -dimensional vector of the weight of the different attribute levels. In our two surveys, $q = 11$ and \mathbf{x}_{ksn} contains the values for all 11 dummy variables for the chosen profile. For instance, for a match in October on Wednesday 20:30 against a top level opponent, $\mathbf{x}'_{ksn} = (0, 1, 0, 1, 0, 0, 0, 0, 0, -1, -1)$. The random error terms ϵ_{ksn} are assumed to be independently and identically type I extreme value distributed. With this assumption on the error term, the probability that respondent n chooses

alternative k between K alternatives in choice set s can be expressed as follows:

$$p_{ksn} = \frac{\exp(\mathbf{x}'_{ksn}\boldsymbol{\beta})}{\sum_{j=1}^K \exp(\mathbf{x}'_{j sn}\boldsymbol{\beta})}. \quad (2)$$

In what follows, we drop the index n in the probabilities as all respondents get the same choice sets. The likelihood of the choices of N respondents in all S choice sets is given by

$$L(\mathbf{y}_N|\mathbf{X}, \boldsymbol{\beta}) = \prod_{n=1}^N \prod_{s=1}^S \prod_{k=1}^K (p_{ks})^{y_{ksn}}, \quad (3)$$

where y_{ksn} is a binary variable that equals 1 if respondent n chooses alternative k in choice set s , and 0 otherwise. The vector $\boldsymbol{\beta}$ which maximizes this likelihood gives the estimated parameter values.

Let $p_s = (p_{1s}, p_{2s}, \dots, p_{Ks})'$, let P_s be a diagonal matrix with diagonal elements p_s , and let \mathbf{X} be the design matrix consisting of the stacked matrices \mathbf{X}_s which consist themselves of all \mathbf{x}_{ks} vectors corresponding to choice set s . The Fisher information matrix can be written in the following closed form

$$I_{FIM}(\boldsymbol{\beta}|\mathbf{X}) = -E\left(\frac{\partial^2 \log L}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}'}\right) = \sum_{s=1}^S \mathbf{X}'_s (P_s - p_s p_s') \mathbf{X}_s. \quad (4)$$

The Fisher information matrix is inversely proportional to the variance-covariance matrix of the parameters, which forms the basis of the D-criterion to be introduced in the following section.

3.3 Semi-Bayesian D-optimal design

The D-criterion is undoubtedly the most frequently used criterion to select the choice sets to be used in a discrete choice experiments. The primary advantage of this criterion over other optimality criteria is that the optimal design is independent of the scale or coding of the attributes (Zwerina et al. 1996; Goos 2002; Kessels et al. 2006; Yu et al. 2009). In line with the previous notation, the D-error is defined as

$$D_{error} = |I_{FIM}(\boldsymbol{\beta}|\mathbf{X})|^{-1/q}. \quad (5)$$

A smaller D-error value indicates a better design. Note that the D-error depends on the design matrix \mathbf{X} and on the parameter vector $\boldsymbol{\beta}$.

We prefer to construct a D-optimal design in a semi-Bayesian context. We take the uncertainty about the prior parameter values into account by minimizing the expected value of the D-error over a prior distribution $\pi(\boldsymbol{\beta})$ of likely parameter values, giving rise to the semi-Bayesian D-error:

$$D_{SB}(\boldsymbol{\beta}, \mathbf{X}) = \int_{\boldsymbol{\beta}} \left| \sum_{s=1}^S \mathbf{X}'_s (P_s - p_s p_s') \mathbf{X}_s \right|^{-1/q} \pi(\boldsymbol{\beta}) d\boldsymbol{\beta}. \quad (6)$$

This criterion leads to designs that are efficient for a whole range of likely parameter values. Notice that the integrals that are needed for the evaluation

of the semi-Bayesian D-error are intractable except in a few special cases. In practice, the semi-Bayesian D-error can be approximated by drawing R random vectors β^r from the prior distribution $\pi(\beta)$, then taking the average of the determinants. Like Sandor and Wedel (2001) and Kessels et al. (2006), in the television broadcasting survey we assumed a Gaussian prior distribution $\pi(\beta)$ with mean $(-0.15, 0.07, -0.01, 0.08, 0.16, -0.24, 0.01, -0.15, 0.24, -0.59, -0.13)'$, and in the stadium attendance survey, with mean $(0.21, 0.04, -0.15, -0.37, 0.24, 0.30, 0.35, 0.25, 0.00, -0.47, -0.07, 0.43, 0.04, 0.09)'$. These prior mean vectors were obtained from the pilot study. The television viewers and stadium spectators seem to have rather different preferences on several attribute levels. However, these mean values were estimated on a small sample (60 respondents), therefore we used a relatively large prior variance of 1 for each parameter. With large prior variances, the design will be relatively efficient for a larger region around the mean values (Zwerina et al. 1996; Sandor & Wedel 2001; Kessels et al. 2006; Yu et al. 2009).

To search for a semi-Bayesian D-optimal design, we generated a random starting design which was iteratively improved by exchanging choice sets from the candidate sets. Each change was followed by a new evaluation of the semi-Bayesian D-error, until no further improvement was found. We repeated this procedure for various starting designs, in order to decrease the risk of ending up with a local minimum.

The table in appendix A gives the best semi-Bayesian D-optimal design matrix found for the television broadcasting survey. The best semi-Bayesian D-optimal design matrix found for the stadium attendance survey is presented in appendix B.

3.4 Modeling heterogeneity in the population

As mentioned above, the conditional logit model assumes that all the respondents attach the same value to an attribute level. This model is therefore estimating the partworths of an “average” respondent. However, if we are not only interested in the average preferences, but also in the individual level preferences, we can fit the mixed logit model.

Let β_n be the q -dimensional coefficient vector representing respondent n 's preferences. In order to catch the heterogeneity among respondents, we assume that β_n follows a multivariate normal distribution $f(\beta_n|\mu_\beta, \Sigma_\beta)$, with mean μ_β and covariance matrix Σ_β . In this case, the main interest lies in the mean vector μ_β and the vector σ_β , which contains the diagonal elements of the heterogeneity matrix Σ_β . The unconditional probability of respondent n 's choice sequence is integrated over the prior distribution of β_n as follows

$$\pi(\mathbf{y}_n|\mathbf{X}, \mu_\beta, \Sigma_\beta) = \int \left(\prod_{s=1}^S \prod_{k=1}^K (p_{ksn})^{y_{ksn}} \right) f(\beta_n|\mu_\beta, \Sigma_\beta) d\beta_n. \quad (7)$$

The corresponding likelihood of the observed choices for N respondents is

$$L(\mathbf{y}_N|\mathbf{X}, \mu_\beta, \Sigma_\beta) = \prod_{n=1}^N \int \left(\prod_{s=1}^S \prod_{k=1}^K (p_{ksn})^{y_{ksn}} \right) f(\beta_n|\mu_\beta, \Sigma_\beta) d\beta_n. \quad (8)$$

The parameter vector μ_β and the matrix Σ_β that maximize the likelihood can be used as point estimates. However, as the integral over the prior dis-

tribution does not have a closed form, maximizing the likelihood function is computationally intractable. To overcome this difficulty, we use Hierarchical Bayes estimation for the mixed logit model. In a Bayesian approach prior information on the parameters (which can be uninformative) is updated with the data to get the posterior distribution of the parameters. The output of such a procedure therefore consists of a sample of the posterior distribution of all the parameters, in this case of μ_{β} , Σ_{β} and β_n . The mean and standard deviation of these posterior distributions can be used as point and standard error estimates for the corresponding parameter. For more details about Hierarchical Bayes estimation, refer to Train (2003) and Orme (2000).

3.5 Data collection

The final version of the questionnaire was distributed online on a national scale, with the help of *Telenet*, *Belgacom*, and the fans forums of several teams in the Belgian Pro League 2013–14. The survey started with the questions about television broadcasting. Respondents were asked whether they regularly watched the Belgian Pro League matches on TV. If not, they were redirected to the second part of the survey. Fans who indicated that they never went to a stadium to watch a Belgian Pro League match were not allowed to fill out the corresponding questions. To increase the response rate, respondents could take part in a lottery for free stadium tickets if they left their email address. Over a 4-week period, we collected 588 responses in all. We retained only the 366 complete responses in the television broadcasting survey and the 307 complete responses in the stadium attendance survey.

4 Results

4.1 Television broadcasting results

4.1.1 Conditional logit estimates

We first fit a conditional logit model to the choice data concerning television broadcasting. Besides the main effects model, we have also estimated an interaction effects model, hoping to detect some interaction effects between kickoff times and opponents. As only the interaction effects between either Wednesday 20:30 or Saturday 20:00 and a middle or lower level opponent are significant, we do not report these results in more detail. Table 4 summarizes the estimated parameters in the main effects model with their standard errors in parentheses. The estimates that are statistically significantly different from zero at the 0.05 level are highlighted in bold. As all the attributes were effects-coded, a significant estimate in this context indicates that it is significantly different from the average effect of that attribute.⁵

This table can be used to predict the utility of a match, given the month, kickoff time and opponent level. For instance, for a match in October on Wednesday 20:30 against a lower level opponent, the predicted utility is $U = 0.083 - 0.299 - 0.610 = -0.826$.

⁵Remark that the parameter and standard error estimates for the reference categories have been derived from the other parameter estimates and the covariance matrix of these estimates.

Table 4: Conditional logit estimates for television broadcasting with standard errors in parentheses.

Month	Kickoff time	Opponent
August	Wednesday 20:30	Lower level opponent
-0.174 (0.024)	-0.299 (0.034)	-0.610 (0.023)
October	Friday 20:30	Middle level opponent
0.083 (0.025)	0.034 (0.033)	-0.041 (0.019)
January	Saturday 18:00	Top level opponent
0.047 (0.026)	-0.174 (0.035)	0.651 (0.025)
March	Saturday 20:00	
0.044 (0.028)	0.383 (0.033)	
	Sunday 14:30	
	-0.103 (0.034)	
	Sunday 18:00	
	0.249 (0.034)	
	Sunday 20:30	
	-0.089 (0.042)	

The conditional logit estimates describe the general preference patterns of Belgian soccer fans. As this sample is a mixture of the fans of all 16 teams, it can be regarded as an average fan. It turns out that August is not a popular month for soccer fans. This seems reasonable, as a considerable proportion of people take holidays in August. October can attract Belgian fans' attention as after several matches in the beginning stage of a season, teams can usually determine their most suitable lineups as well as the most appropriate tactics, leading to a more stable performance in October. The fans tend to have an average attitude towards matches in January and March. It is not surprising that the ending phase of the regular season is not particularly important, as there will be play-offs soon after. People on average dislike matches scheduled for Wednesday 20:30, Saturday 18:00, Sunday 14:30, or Sunday 20:30. Instead, they like to watch matches on Saturday 20:00 and Sunday 18:00. They have an average preference for matches on Friday 20:30. These patterns are in accordance with our expectation that the Belgian people on average have an aversion to watching evening matches on the eve of a working day, and usually have alternative options during the daytime in the weekend. Not surprisingly, most of the fans have no interest in watching a match against a lower level opponent. They have a small negative attitude towards a middle level opponent, too. As expected, fans are interested in watching a match against a top level opponent. To examine the heterogeneity among the respondents, we will fit also a mixed logit model to the choice data.

4.1.2 Mixed logit estimates

To capture the individual level preferences, we use a Hierarchical Bayes approach in the R-package *Bayesm* to estimate the mixed logit model regarding the television broadcasting data. We specified the prior distribution of the mean μ_{β} as a Gaussian distribution $N(0_{11}, I_{11})$. The prior mean of the covariance matrix Σ_{β} is assumed to be an 11-dimensional identity matrix. The covariance matrix of the prior for the covariances is assumed to follow an inverse-Wishart distribution with 11 degrees of freedom and scale matrix I_{11} . In Table 5 we

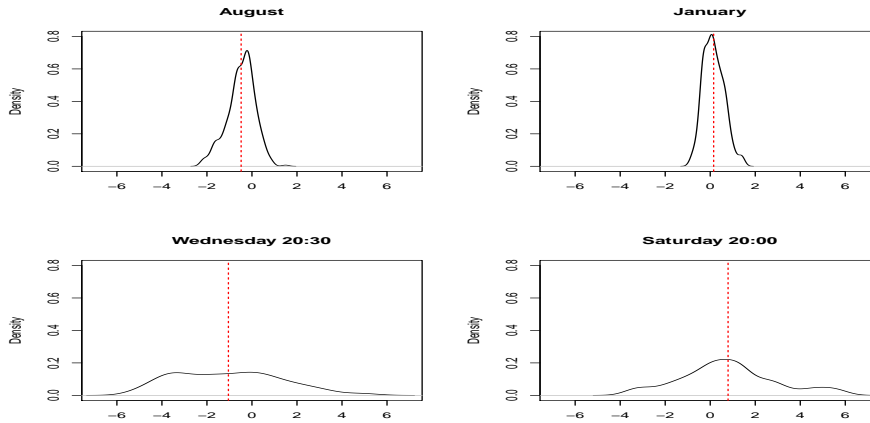


Figure 1: Density curves of the posterior means of the individual parameters β_n for August, January, Wednesday 20:30 and Saturday 20:00. The mean is highlighted by the red dashed line. Results for television broadcasting.

level opponents. As most of the 11 mixed logit estimates are not significantly different between the three groups, we do not report these results.

4.1.3 Predictions & validation

In the previous section we have used both the conditional logit model and the mixed logit model to estimate the parameters. By using each set of estimates, an average fan’s expected utility of watching a Belgian Pro League match against a top, middle, or lower level opponent can be calculated for each month and kickoff time combination. Table 6 displays the predicted utilities by using the estimates from the two models. As can be seen from Table 6, the predicted utilities using the mixed logit estimates are generally larger in absolute value than those using the conditional logit estimates, which is in accordance with the larger estimates (Tables 4 & 5). However, the correlation between the conditional logit utilities and the mixed logit utilities is 0.99.

To validate these predictions, we compared them to observed audience rating data. *Telenet* has broadcast all the regular stage matches of the Belgian Pro League 2013–14. They provided us with a dataset containing relative television ratings of all these matches.⁶ The match with the highest recorded rating is given a benchmark value of 100. Then all other matches’ ratings are given as a percentage of the maximum rating. As these are overall ratings recorded from all the viewers in Belgium, they can be regarded as a measure of an average Belgian soccer fan’s preference for watching a match. Therefore, these data can be used to examine the accuracy of our predictions.

In order to make the TV ratings comparable to our predicted utilities, we classify the TV data into the 28 month and kickoff time combinations specified in our experiment. August, October, January and March are retained as four

⁶For confidentiality reasons, the raw data were not provided. This poses no problem as the results of a discrete choice experiments are also to be interpreted in a relative way.

Table 6: Predicted utilities of watching Belgian Pro League matches under all combinations of month, kickoff time, and opponent level, using either the conditional logit (CL) estimates or the mixed logit (ML) estimates.

Setting	Top opponent		Middle opponent		Lower opponent	
	CL	ML	CL	ML	CL	ML
AUG WED 20:30	0.177	0.060	-0.514	-1.529	-1.083	-3.112
AUG FRI 20:30	0.511	1.361	-0.181	-0.228	-0.750	-1.811
AUG SAT 18:00	0.303	1.001	-0.388	-0.588	-0.958	-2.172
AUG SAT 20:00	0.860	1.900	0.168	0.311	-0.401	-1.273
AUG SUN 14:30	0.374	0.798	-0.318	-0.791	-0.887	-2.374
AUG SUN 18:00	0.726	1.719	0.035	0.130	-0.535	-1.454
AUG SUN 20:30	0.388	0.908	-0.303	-0.681	-0.873	-2.265
OCT WED 20:30	0.434	0.743	-0.257	-0.847	-0.826	-2.430
OCT FRI 20:30	0.768	2.043	0.076	0.454	-0.493	-1.129
OCT SAT 18:00	0.560	1.683	-0.131	0.094	-0.701	-1.490
OCT SAT 20:00	1.117	2.582	0.425	0.993	-0.144	-0.591
OCT SUN 14:30	0.631	1.480	-0.061	-0.109	-0.630	-1.692
OCT SUN 18:00	0.983	2.401	0.292	0.812	-0.278	-0.772
OCT SUN 20:30	0.645	1.590	-0.046	0.001	-0.616	-1.582
JAN WED 20:30	0.398	0.693	-0.293	-0.896	-0.862	-2.479
JAN FRI 20:30	0.732	1.994	0.040	0.405	-0.529	-1.178
JAN SAT 18:00	0.524	1.634	-0.167	0.045	-0.737	-1.539
JAN SAT 20:00	1.081	2.533	0.389	0.944	-0.180	-0.640
JAN SUN 14:30	0.595	1.431	-0.097	-0.158	-0.666	-1.741
JAN SUN 18:00	0.947	2.352	0.256	0.763	-0.314	-0.821
JAN SUN 20:30	0.609	1.541	-0.082	-0.048	-0.652	-1.632
MAR WED 20:30	0.395	0.667	-0.297	-0.922	-0.866	-2.505
MAR FRI 20:30	0.728	1.968	0.036	0.379	-0.533	-1.204
MAR SAT 18:00	0.520	1.608	-0.171	0.019	-0.741	-1.565
MAR SAT 20:00	1.077	2.507	0.386	0.918	-0.184	-0.666
MAR SUN 14:30	0.591	1.405	-0.101	-0.184	-0.670	-1.767
MAR SUN 18:00	0.943	2.326	0.252	0.737	-0.318	-0.847
MAR SUN 20:30	0.605	1.515	-0.086	-0.074	-0.656	-1.658

typical representatives for the first, second, third and fourth quartiles of a regular season. Every candidate matchday is tied to the nearest month in our survey. For instance, September 20th should be classified under October. Similarly, we retain the 7 kickoff times specified in the experimental design. Beyond these routine kickoff times, Thursday 14:30, Thursday 18:00, and Thursday 20:30 have appeared at least once each in the TV ratings data. We will tie these observations either to Wednesday (due to double matchdays in one week) or to Sunday (due to Christmas holiday). The top–middle–lower division of the 16 teams is retained to approximate the quality of a match. In this way every match can be put into one of the following classes: top vs. top, top vs. middle, top vs. lower, middle vs. middle, middle vs. lower, and lower vs. lower.

Table 7 presents the average relative TV ratings of the matches classified according to the month, the kickoff time, and the paired opponents. The number of observations is given in parentheses. For more than half (87 out of 168) of the elements in this table, there are no observations at all, as many month and kickoff time combinations have never been used for matches for certain opponent-pair levels. The elements of Table 7 reflect the relative interest of an average fan in watching the Belgian Pro League matches for six opponent-pair levels, while the elements of Table 6 describe an average fan’s interest in

watching a match of his favorite team against either a top (t), middle (m), or lower (l) level opponent. Therefore, these two tables cannot be directly compared.

Table 7: Average relative TV ratings.

Setting	t vs. t	t vs. m	t vs. l	m vs. m	m vs. l	l vs. l
AUG WED 20:30	-	-	-	-	-	-
AUG FRI 20:30	-	-	52 (2)	32 (1)	-	-
AUG SAT 18:00	-	44 (2)	-	-	30 (1)	-
AUG SAT 20:00	-	44 (1)	21.67 (3)	14 (3)	9.42 (12)	4 (10)
AUG SUN 14:30	-	51 (3)	36.5 (2)	42 (1)	-	-
AUG SUN 18:00	-	64.25 (4)	38 (4)	-	8 (1)	-
AUG SUN 20:30	78 (1)	55 (1)	45.25 (4)	-	-	-
OCT WED 20:30	-	20 (1)	37.67 (3)	10 (1)	4 (2)	2 (1)
OCT FRI 20:30	-	-	46 (1)	28 (1)	34 (1)	-
OCT SAT 18:00	-	-	45 (2)	39 (1)	-	-
OCT SAT 20:00	-	15 (1)	27.67 (6)	14 (4)	7.61 (18)	3.55 (11)
OCT SUN 14:30	70 (3)	51.5 (2)	39.5 (2)	-	-	-
OCT SUN 18:00	76.5 (2)	66 (5)	41 (1)	-	-	-
OCT SUN 20:30	-	55 (1)	32.5 (2)	-	22.67 (3)	35 (1)
JAN WED 20:30	-	-	-	-	-	-
JAN FRI 20:30	-	43.67 (3)	57 (1)	-	-	-
JAN SAT 18:00	-	58 (1)	41.75 (4)	-	-	-
JAN SAT 20:00	-	24.5 (2)	33 (2)	12.67 (3)	8.67 (15)	3.75 (12)
JAN SUN 14:30	55 (1)	50.25 (4)	26.67 (3)	-	3.67 (3)	2 (1)
JAN SUN 18:00	88 (3)	55.25 (4)	63 (1)	46 (1)	-	-
JAN SUN 20:30	-	-	50.33 (3)	34 (1)	9 (1)	-
MAR WED 20:30	-	-	-	-	-	-
MAR FRI 20:30	-	64 (1)	-	37 (1)	-	-
MAR SAT 18:00	-	-	33 (1)	-	-	-
MAR SAT 20:00	-	19 (2)	15 (3)	-	7.09 (11)	5.4 (5)
MAR SUN 14:30	67 (1)	-	41 (1)	44 (1)	-	-
MAR SUN 18:00	77 (1)	37 (2)	27.75 (4)	11 (1)	3 (2)	3 (1)
MAR SUN 20:30	-	-	41 (1)	-	-	-

We make use of the average utilities given the opponent level in Table 6 to approximate an average fan's utility from watching any Belgian Pro League match. Given a specified month and a kickoff time, let v_i denote the predicted utility an average fan obtains from watching a match between their favorite team and a level i team ($i = t, m, l$), u_{ij} denoting the utility an average fan can obtain from watching a match between a level i team against a level j team ($i, j = t, m, l$), we can approximate u_{ij} by v_i and v_j as follows

$$\begin{aligned}
u_{ii} &\approx v_i, \\
u_{ij} &\approx \log(\exp(v_i) + \exp(v_j)).
\end{aligned} \tag{9}$$

The idea behind this approximation is as follows. We have the predicted utilities v_t, v_m, v_l of watching a match against a top, middle or lower level opponent. For matches between two teams of the same level, say top, only one source of information v_t is available. Whereas for matches between two teams of different competitive levels, say top and middle, we have two sources of information: v_t and v_m which can be used to approximate an average fan's utility of watching such a match in a mixed way.

We have also considered several alternative approximations, for instance, using the group-specific estimates to approximate u_{ij} , or to weight them by group size. However, such approximations did not result in better predictions and are therefore not reported here. So in the remainder of this paper we will use formula (9) to predict the utilities.

Table 9 gives the predicted utilities of an average fan in watching all possible Belgian Pro League matches specified by month, kickoff time, and opponent pair, using either the conditional logit estimates or the mixed logit estimates. These utilities can now be compared to Table 7.

We use the Pearson correlation coefficient to check the linear dependence between the predicted utilities (Table 9) and the TV ratings (Table 7). Table 8 summarizes the results and includes the p -values for testing the significance of these results. It turns out that all correlations are significantly different from zero. The first column in Table 8 gives the Pearson correlations based on the 81 non-missing observations in Table 7. The utilities computed with the mixed logit estimates have a slightly higher correlation with the TV ratings than those using the conditional logit estimates.

Table 8: Unweighted and weighted Pearson correlations between the average relative TV ratings and the predicted utilities using either the conditional logit estimates or the mixed logit estimates.

Model	Weights		
	1	sample size	$\frac{1}{\sigma^2}$
Conditional logit model	0.543 (p<e-5)	0.587 (p<e-5)	0.965 (p<e-5)
Mixed logit model	0.587 (p<e-5)	0.646 (p<e-5)	0.965 (p<e-5)

We constructed a bubble plot (Figure 2) to illustrate the correlation between the TV ratings and the predicted utilities using the mixed logit estimates. The size of each bubble in the plot is proportional to its sample size.

As the 81 bubbles on Figure 2 have considerably diverse sample sizes, it is reasonable to take the relative importance of each bubble into account when calculating the correlation. To this end, we also report the correlation weighted with the sample size in the second column of Table 8. The regression line is obtained by weighted least squares with weights equal to the number of observations.

Some bubbles do not lie close to the regression line. Especially, some bubbles concerning Saturday 20:00 in the upper-left corner of the figure, seem to be overrated. However, this overrate is debatable, as Saturday 20:00 traditionally is the only slot at which four or even five matches are simultaneously played, leading to a considerable split flow in audience ratings. Most of the other outliers on Figure 2 might be due to either combinations containing quite few matches, or those with very low ratings.

Furthermore, some of the average ratings in Table 7 are more reliable than others, as some are based on similar relative TV ratings (e.g., there are 11 “lower vs. lower” matches on Saturday 20:00 in October with ratings: 5, 3, 3, 2, 7, 6, 3, 2, 3, 2, 3) whereas others are based on very different relative TV ratings (e.g., there are two “top vs. middle” matches on Sunday 18:00 in March with ratings: 15 and 59). To this end, we examined the variance of observations

in each bubble. The last column of Table 8 gives the correlations weighted with the inverse of the variance of the ratings in each bubble, where only 42 bubbles with at least two observations are retained. Again, the mixed logit estimates outperform the conditional logit model estimates in this measure. We can conclude that there is a high correlation between our predicted utilities and the TV ratings, validating the use of our model to predict audience ratings.

Table 9: Predicted utilities of watching Belgian Pro League matches for all combinations of month, kickoff time, and opponent-pair, using either the conditional logit (CL) estimates or the mixed logit (ML) estimates.

Setting	t vs. t		t vs. m		t vs. l		m vs. m		m vs. l		l vs. l	
	CL	ML	CL	ML	CL	ML	CL	ML	CL	ML	CL	ML
AUG WED 20:30	0.177	0.060	-0.110	-0.447	-0.266	-0.592	-0.514	-1.529	-0.759	-2.035	-1.083	-3.112
AUG FRI 20:30	0.511	1.361	0.223	0.854	0.067	0.709	-0.181	-0.228	-0.426	-0.734	-0.750	-1.811
AUG SAT 18:00	0.303	1.001	0.016	0.493	-0.141	0.349	-0.388	-0.588	-0.633	-1.095	-0.958	-2.172
AUG SAT 20:00	0.860	1.900	0.573	1.392	0.416	1.248	0.168	0.311	-0.076	-0.196	-0.401	-1.273
AUG SUN 14:30	0.374	0.798	0.086	0.291	-0.070	0.146	-0.318	-0.791	-0.563	-1.297	-0.887	-2.374
AUG SUN 18:00	0.726	1.719	0.439	1.211	0.282	1.067	0.035	0.130	-0.210	-0.377	-0.535	-1.454
AUG SUN 20:30	0.388	0.908	0.101	0.400	-0.056	0.256	-0.303	-0.681	-0.548	-1.188	-0.873	-2.265
OCT WED 20:30	0.434	0.743	0.147	0.235	-0.009	0.090	-0.257	-0.847	-0.502	-1.353	-0.826	-2.430
OCT FRI 20:30	0.768	2.043	0.481	1.536	0.324	1.391	0.076	0.454	-0.168	-0.052	-0.493	-1.129
OCT SAT 18:00	0.560	1.683	0.273	1.175	0.116	1.031	-0.131	0.094	-0.376	-0.413	-0.701	-1.490
OCT SAT 20:00	1.117	2.582	0.830	2.074	0.673	1.930	0.425	0.993	0.181	0.486	-0.144	-0.591
OCT SUN 14:30	0.631	1.480	0.344	0.973	0.187	0.828	-0.061	-0.109	-0.305	-0.615	-0.630	-1.692
OCT SUN 18:00	0.983	2.401	0.696	1.893	0.539	1.749	0.292	0.812	0.047	0.305	-0.278	-0.772
OCT SUN 20:30	0.645	1.590	0.358	1.083	0.201	0.938	-0.046	0.001	-0.291	-0.506	-0.616	-1.582
JAN WED 20:30	0.398	0.693	0.111	0.186	-0.045	0.041	-0.293	-0.896	-0.538	-1.402	-0.862	-2.479
JAN FRI 20:30	0.732	1.994	0.445	1.487	0.288	1.342	0.040	0.405	-0.204	-0.101	-0.529	-1.178
JAN SAT 18:00	0.524	1.634	0.237	1.126	0.080	0.982	-0.167	0.045	-0.412	-0.462	-0.737	-1.539
JAN SAT 20:00	1.081	2.533	0.794	2.025	0.637	1.881	0.389	0.944	0.145	0.437	-0.180	-0.640
JAN SUN 14:30	0.595	1.431	0.308	0.924	0.151	0.779	-0.097	-0.158	-0.341	-0.664	-0.666	-1.741
JAN SUN 18:00	0.947	2.352	0.660	1.844	0.503	1.700	0.256	0.763	0.011	0.256	-0.314	-0.821
JAN SUN 20:30	0.609	1.541	0.322	1.033	0.165	0.889	-0.082	-0.048	-0.327	-0.555	-0.652	-1.632
MAR WED 20:30	0.395	0.667	0.108	0.160	-0.049	0.015	-0.297	-0.922	-0.541	-1.428	-0.866	-2.505
MAR FRI 20:30	0.728	1.968	0.441	1.461	0.284	1.316	0.036	0.379	-0.208	-0.127	-0.533	-1.204
MAR SAT 18:00	0.520	1.608	0.233	1.100	0.077	0.955	-0.171	0.018	-0.416	-0.488	-0.741	-1.565
MAR SAT 20:00	1.077	2.507	0.790	1.999	0.633	1.854	0.386	0.918	0.141	0.411	-0.184	-0.666
MAR SUN 14:30	0.591	1.405	0.304	0.898	0.147	0.753	-0.101	-0.184	-0.345	-0.690	-0.670	-1.767
MAR SUN 18:00	0.943	2.326	0.656	1.818	0.500	1.673	0.252	0.737	0.007	0.230	-0.318	-0.847
MAR SUN 20:30	0.605	1.515	0.318	1.007	0.162	0.863	-0.086	-0.074	-0.331	-0.581	-0.656	-1.658

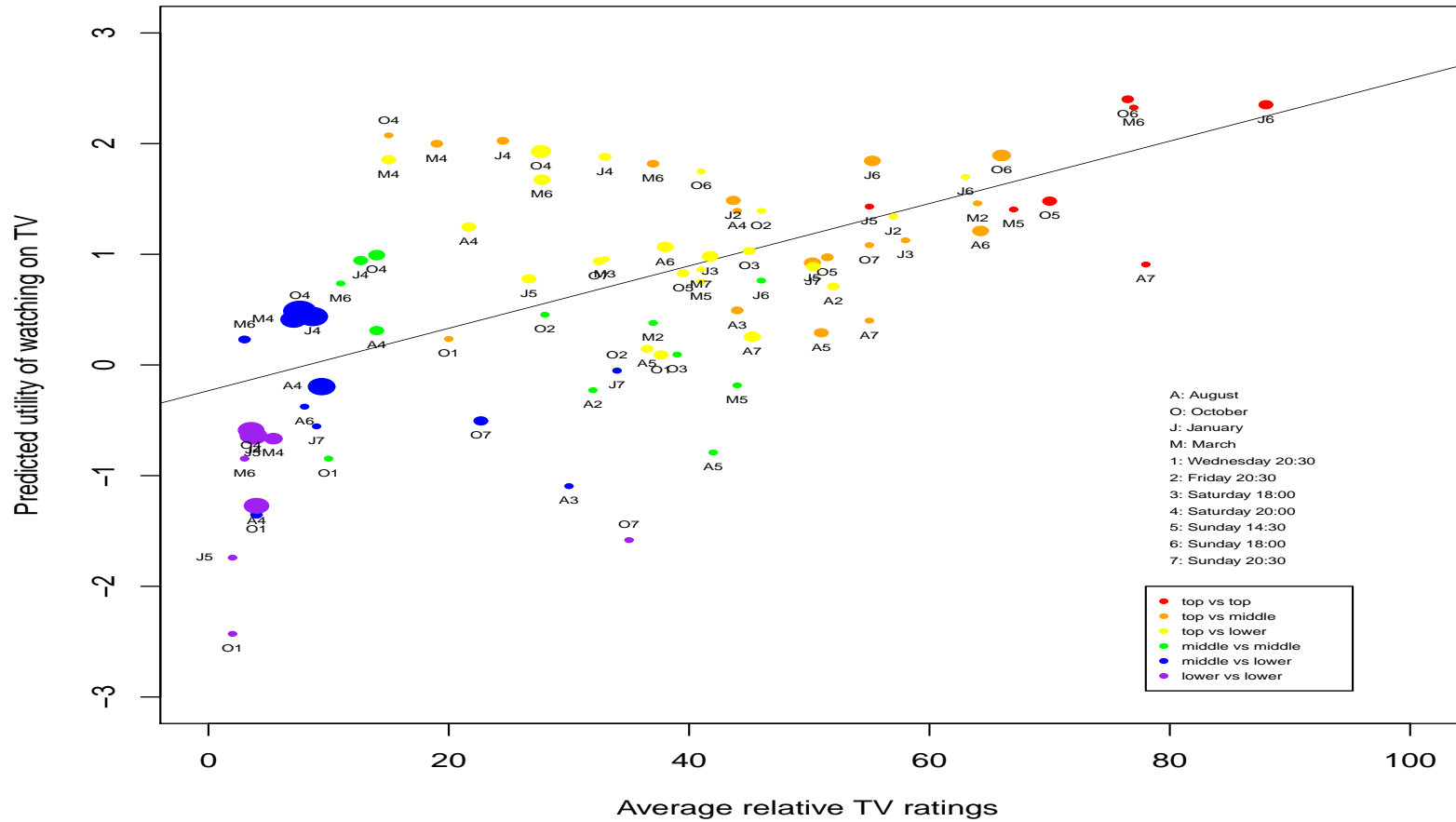


Figure 2: Bubble plot of the average relative TV ratings and the predicted utilities based on the mixed logit estimates.

4.2 Stadium attendance results

As introduced in Section 1, Belgian clubs prefer a schedule that attracts as many fans as possible to their home matches. To capture the preferences of the home crowd with respect to the schedule, we also conducted a stadium attendance survey. We used the attributes month, kickoff time, and opponent, for parameter estimation and predictions.

4.2.1 Conditional logit estimates

In this section, we briefly summarize the estimates using the conditional logit model. Table 10 gives the conditional logit estimates (standard errors in parentheses). Eleven out of the 14 parameters are significantly different from the average effect of the corresponding attribute at the 0.05 level.

Comparing Table 4 and 10 one can find some striking differences with respect to the preferences for the different months and kickoff times. For instance, now fans do not like January, presumably as it is then cold outside. Perhaps due to spring fever, March is rather popular. Apparently, August’s nice summer weather is neutralized by the fact that many people leave on holiday. Wednesday 20:30 and Sunday 20:30 are not good slots for stadium attendance, as the next day is a working day. In contrast, Saturday 18:00 and 20:00, and Sunday 14:30 and 18:00 are preferred for stadium attendance. The fans have an average preference for Friday 20:30. No surprise, fans like to attend a match against a top level opponent, but are reluctant to travel to the stadium for lower or even middle level opponents.

Table 10: Conditional logit estimates for stadium attendance with standard errors in parentheses.

Month	Kickoff time	Opponent
August	Wednesday 20:30	lower level opponent
0.023 (0.032)	-0.782 (0.051)	-0.620 (0.046)
October	Friday 20:30	Middle level opponent
0.018 (0.040)	0.064 (0.043)	-0.111 (0.036)
January	Saturday 18:00	Top level opponent
-0.220 (0.045)	0.158 (0.070)	0.731 (0.057)
March	Saturday 20:00	
0.179 (0.055)	0.772 (0.047)	
	Sunday 14:30	
	0.148 (0.060)	
	Sunday 18:00	
	0.210 (0.061)	
	Sunday 20:30	
	-0.570 (0.073)	

4.2.2 Mixed logit estimates

The mixed logit estimates obtained from the R-package *bayesm* are summarized in Table 11, similarly as was done for the TV broadcasting results. Similar patterns of the estimates can be observed as in Table 10. To illustrate the degree of heterogeneity, we plotted the density curves of the 307 means of the posterior

distributions of the individual parameters β_n of the parameters for August, January, Wednesday 20:30 and Saturday 20:00 (Figure 3). Again, the heterogeneity for the parameters for Wednesday 20:30 or Saturday 20:00 is clearly larger than that of August or January. A higher preference for August and a higher aversion for January are observed in case of stadium attendance. In contrast, the stadium fans have more similar preferences for matches on Wednesday 20:30 or Saturday 20:00 than the TV viewers. They show a clearly higher preference for Saturday 20:00 and also a higher aversion to Wednesday 20:30 than do TV viewers.

Table 11: Mixed logit estimates for stadium attendance: the means and standard deviations of the posterior distributions of μ_β and σ_β .

μ_β (s.e.)	σ_β (s.e.)	μ_β (s.e.)	σ_β (s.e.)	μ_β (s.e.)	σ_β (s.e.)
Month		Kickoff time		Opponent	
	August	Wednesday 20:30		Lower level opponent	
-0.052 (0.085)	1.041 (0.084)	-1.670 (0.146)	1.724 (0.151)	-1.437 (0.130)	1.442 (0.149)
	October	Friday 20:30		Middle level opponent	
0.020 (0.075)	0.664 (0.059)	0.162 (0.106)	1.380 (0.116)	-0.300 (0.088)	0.878 (0.088)
	January	Saturday 18:00		Top level opponent	
-0.422 (0.099)	0.895 (0.093)	0.324 (0.129)	1.062 (0.138)	1.737 (0.171)	
	March	Saturday 20:00			
0.454 (0.116)		1.701 (0.163)	2.109 (0.173)		
		Sunday 14:30			
		0.380 (0.139)	1.706 (0.163)		
		Sunday 18:00			
		0.366 (0.137)	1.423 (0.148)		
		Sunday 20:30			
		-1.263 (0.174)			

4.2.3 Predictions & validation

Similar to the case of television broadcasting, we can use these estimates to predict the utilities an average fan obtains in watching Belgian Pro League home matches of his favorite team. To validate our predictions, we compared them to the stadium attendance in all the 240 regular stage matches in Belgian Pro League 2013–14. The attendances in this data strictly refer to the home team fans, classified in the same way as for television broadcasting. Table 12 summarizes the three versions of Pearson correlations with corresponding p -values for the test of their significance between the average observed home attendances and the predicted utilities using either the conditional or mixed logit estimates. We constructed a bubble plot (Figure 4) to illustrate the linear correlation between the observed home attendances and the predicted utilities using the mixed logit estimates. The regression line is obtained by weighted least squares with weights equal to the number of observations. The size of each bubble in the plot is again proportional to its sample size. As expected, the predicted utilities are now less correlated with the observed values than in the TV broadcasting case, as there are many more factors that can influence a fan’s decision whether or not to travel to the stadium, than there are for watching a match on TV.

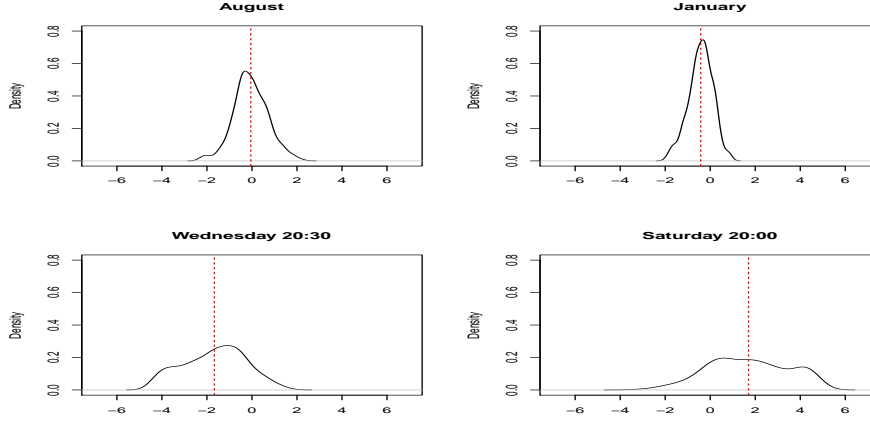


Figure 3: Density curves of the posterior means of the individual parameters β_n for August, January, Wednesday 20:30 and Saturday 20:00. The mean is highlighted by the red dashed line. Results for stadium attendance.

Table 12: Unweighted and weighted Pearson correlations between the observed home attendances and the predicted utilities using either the conditional logit estimates or the mixed logit estimates.

Model	Weights		
	1	sample size	$\frac{1}{\sigma^2}$
Conditional logit model	0.411 (p=1.382e-4)	0.362 (p<e-4)	0.853 (p<e-4)
Mixed logit model	0.448 (p=2.776e-5)	0.456 (p<e-5)	0.856 (p<e-5)

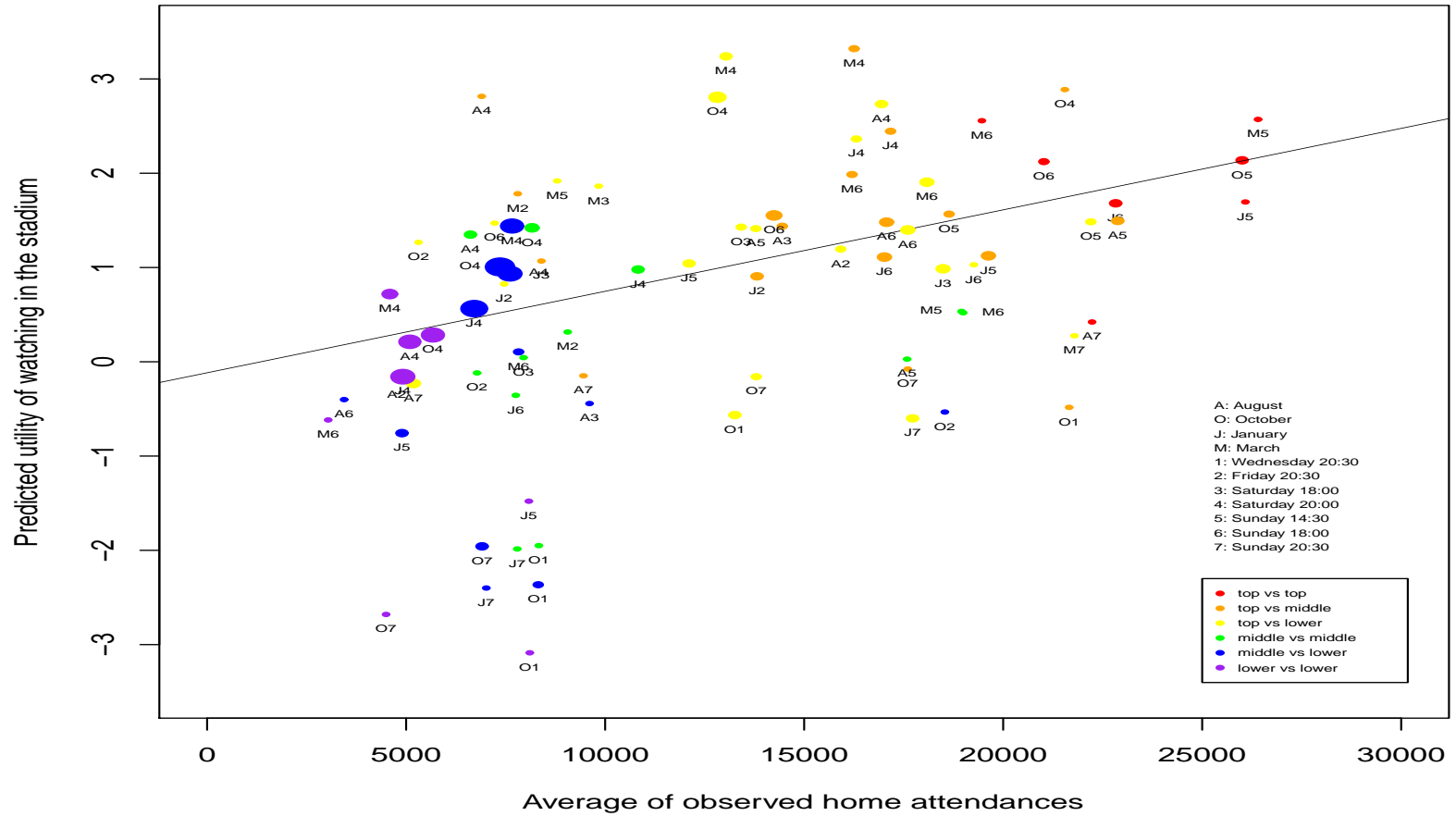


Figure 4: Bubble plot of the observed home attendance and the predicted utilities.

4.3 Similarity between TV viewers and home attendances

As stated in the Introduction, television companies seek a schedule that maximizes the average audience ratings. The clubs also want a schedule that increases the average attendance for their home matches. We will assess to what extent maximizing audience rating and stadium attendance can go together, by examining fans' preferences for the month, the kickoff time, and the opponent level for each match. As the mixed logit utilities in both experiments have the highest correlation with the observed data, we retain the two sets of the mixed logit posterior means of μ_{β} for further analysis (Tables 5 & 11).

Figure 5 illustrates the relation between the predicted utilities of watching a match on TV and in a stadium. The regression line is weighted with the number of observations. The regression line is again obtained by weighted least squares with weights equal to the number of observations. The bubbles in the figure correspond to the combinations that have been observed at least once in the Belgian Pro League 2013–14. The size of each bubble is again proportional to its sample size. The triangles correspond to the combinations that have not been used. The Pearson correlation coefficient between these two complete sets of utilities is 0.866. When applied to the two subsets of utilities with only observed combinations, the Pearson correlation is 0.816, and is 0.750 weighted with the number of observations. A large proportion of the bubbles are close to the regression line. The bubbles referring to matches scheduled on Saturday 20:00 are mostly distributed above the regression line, indicating that stadium attendance is favored. A large proportion of the bubbles regarding matches on Sunday 14:30 is observed above the line. Most of the bubbles concerning August or March are also above the line, confirming the seasonal preference. In contrast, January matches are definitely more preferred to be watched on TV due to the coldness outside. The bubbles concerning matches on Sunday 20:30 are all below the line, which might be due to the inconvenience of attending a match in a stadium that late before a working day. Notice that there are triangles with high predicted utilities in both aspects, e.g., top matches on Saturday 20:00 in either August, October, January or March. These options have not been used for scheduling the fixtures so far, but could clearly be beneficial to clubs as well as to TV broadcasters.

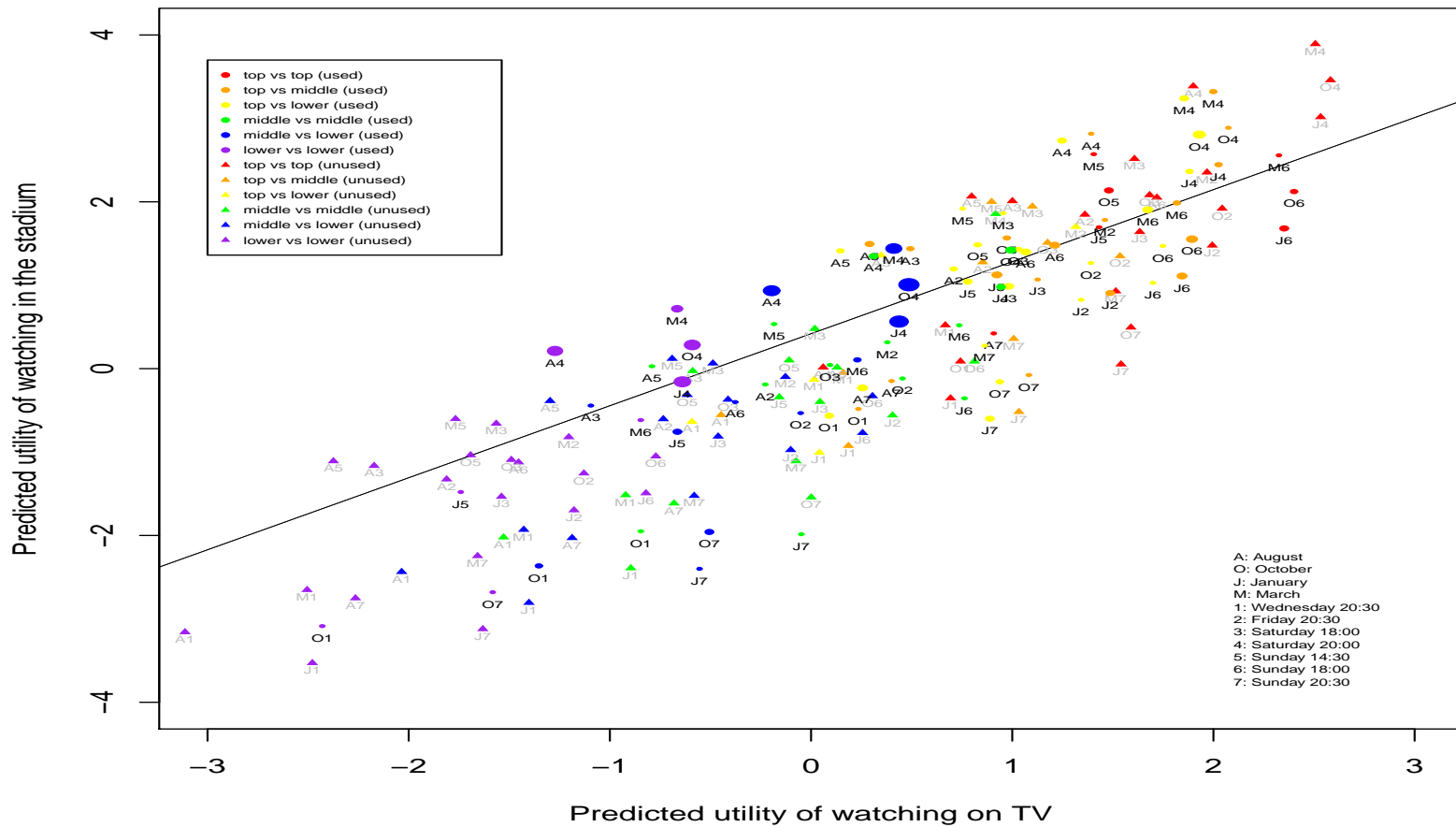


Figure 5: Plot of the predicted utilities of watching matches on TV and in a stadium. Bubbles denote the combinations that have been observed at least once in the Belgian Pro League 2013–14.

5 Towards a schedule that attracts more fans

In this section, we show how our results can be used to improve the scheduling process of the Belgian Pro League in order to persuade more fans to watch the matches on TV.

For the first step in creating the schedule, the Pro League decides on the dates on which rounds need to be played. Our results (see Tables 4 & 5) suggest that matches on Wednesdays are to be avoided at all cost. However, if there are more rounds than available weekends, a midweek matchday is best planned in October or March to ensure a reasonably sized audience. A winter break is reasonable not just to provide a rest period for the players, but also since this period is not very popular with the fans. Finally, it does not seem beneficial to start the season early in August, or even late July, as happened frequently in the last years.

Hitherto, the scheduling process has assigned matches to rounds, taking into account various constraints (Goossens & Spieksma 2009). Given this schedule, broadcasters assign matches to kick-off times round per round, according to their preferences. The utilities in Table 9 can be used to predict audience ratings for each (type of) match, depending on the planned month and kick-off time. This allows us to generate a schedule that assigns matches directly to timeslots, optimizing the expected audience ratings over all rounds. Given a schedule, the broadcasting companies can also use the results from our choice experiment to predict the market share to be expected from each of the simultaneously televised live matches. This may prove valuable information when negotiating with advertisers. The same exercise can be made if stadium attendance has to be maximized. Notice that when the Pro League decides on the relative importance of a fan watching a match on TV compared to a fan buying a stadium ticket, this trade-off can easily be taken into account using a sum of weighted predictions as an input to the scheduling process.

Notwithstanding these applications, our work leaves room for future research. For instance, our analysis is based on matches between pairs of opponents classified as top, middle, or lower level teams. Further refinement of these classes, ultimately to individual teams, will allow us to enhance our predictions and fine-tune the optimization of the schedule. Furthermore, our study did not focus on possible interaction effects between matches. For instance, can we expect more viewers when two top matches are scheduled consecutively on the same day than when we schedule these matches on different days? Obviously, this also relates to pricing and marketing decisions taken by the television companies.

6 Conclusion

In this study we aimed to investigate the impact of the schedule on TV viewership and stadium attendance for the Belgian Pro League. To this end, we conducted a discrete choice experiment to capture the relationship between these scheduling attributes (month, kickoff time and opponent level) and Belgian soccer fans' preferences with respect to watching a Pro League match on TV and his preferences with respect to attending it in the stadium. Both a conditional logit model assuming homogenous preferences and a mixed logit model assum-

ing heterogeneous preferences were fitted to the choice data. For both television broadcasting and stadium attendance, the conditional and mixed logit estimates show similar patterns with respect to the mean preferences. With these estimates, the utility Belgian soccer fans obtain in watching their favorite team's league match either on TV or in the home stadium can be predicted, when the month, the kickoff time, and the opponent are all given. To validate these predictions, we compared the predicted broadcasting utilities with audience ratings in the 2013–14 season and found a high correlation between them. The predicted attendance utilities are moderately correlated to the observed home attendance data in the same season. The stadium fans tend to be more sensitive to the season of the year and to matches on the eve of a working day than are television viewers.

Our predicted utility proves a reliable measure to forecast Belgian fans' tendency to watch any Belgian Pro League match, which is valuable input for the scheduling process, allowing schedules that persuade more fans to watch the matches on TV or in a stadium.

References

- Alavy, K., Gaskell, A., Leach, S., Szymanski, S. (2010) On the edge of your seat: Demand for football on television and the uncertainty of outcome hypothesis. *International Journal of Sport Finance* **5**, 2, 75–95.
- Allan, G., Roy, G. (2008) Does television crowd out spectators? New evidence from the Scottish Premier League. *Journal of Sports Economics*, **9**, 592–605.
- Atkinson, A. C., Donev, A. N. (1992) *Optimum Experimental Designs*. Clarendon Press, Oxford.
- Baimbridge, M., Cameron, S., Dawson, P. (1996) Satellite television and the demand for football: A whole new ballgame? *Scottish Journal of Political Economy*, **43**, 317–333.
- Borland J., Macdonald, R. (2003) Demand for sport. *Oxford Review of Economic Policy*, **19**, 4, 478–502.
- Buraimo, B. (2008) Stadium attendance and television audience demand in English league football. *Manage. Decis. Econ.*, **29**, 513–523.
- Carson, R. T., Louviere, J. J., Anderson, D. A., Arabie, P., Bunch, D. S., Hensher, D. A., Johnson, R. M., Kuhfeld, W. F., Steinberg, D., Swait, J., Timmermans, H., J. B. Wiley, J. B. (1994) Experimental analysis of choice. *Marketing Letters*, **5**, 351–368.
- Carvalho, M. D., Scheerder J., Boen F., Sarmento J. P. (2013) What brings people into the soccer stadium? The case of Belgium and Portugal from a marketing perspective. *Sport Policy & Management Report*, **18**, KU Leuven.
- Feddersen, A., Rott, A. (2011) Determinants of demand for televised live football: Features of the German national football team. *Journal of Sports Economics*, **12**, 3, 352–369.
- Forrest, D., Simmons, R., Buraimo, B. (2005) Outcome uncertainty and the couch potato audience. *Scottish Journal of Political Economy*, **52**, 641–661.
- Forrest, D., Simmons, R. (2006) New issues in attendance demand: The case of the English football league *Journal of Sports Economics*, **7**, 3, 247–266.
- Garcia, J., Rodriguez, P. (2002) The determinants of football match attendance revisited: Empirical evidence from the Spanish football league. *Journal of Sports Economics*, **3**, 18–38.
- Goos, P. (2002) *The Optimal Design of Blocked and Split-Plot Experiments*, Springer-Verlag, New York.
- Goossens, D., Spieksma, F. (2009) Scheduling the Belgian soccer league. *Interfaces*, **39**, 2, 109–118.
- Goossens, D., Spieksma, F. (2012) Soccer schedules in Europe: An overview. *Journal of Scheduling*, **15**, 5, 641–651.
- Kessels, R., P. Goos, M. Vandebroek. (2006) A comparison of criteria to design efficient choice experiments. *Journal of Marketing Research*, **43**, 409–419.

- Kessels, R., P. Goos, M. Vandebroek. (2009) An efficient algorithm for constructing Bayesian optimal choice designs. *Journal of Business & Economic Statistics*, **27**, 2, 279–291.
- McFadden, D. (1974) Conditional logit analysis of qualitative choice behavior. In: *Frontiers in Econometrics*. P. Zarembka (ed.), Academic Press, New York.
- Lynch, J. G. (2005) The effort effects of prizes in the second half of tournaments. *Journal of Economic Behavior & Organization*, **57**, 1, 115–129.
- Madalozzo, R., Villar, R. B. (2009) Brazilian football: What brings fans to the game? *Journal of Sports Economics*, **10**, 6, 639–650.
- Orme B. (2000) Hierarchical Bayes: Why all the attention? Sawtooth Software research paper series.
- Peeters, T., Szymanski, S. (2014) Financial fair play in European football. *Economic Policy*, **29**, 78, 343–390.
- Sandor, Z., Wedel, M. (2001) Designing conjoint choice experiments using managers' prior beliefs. *Journal of Marketing Research*, **38**, 430–444.
- Sandor, Z., Wedel, M. (2005) Heterogeneous conjoint choice designs. *Journal of Marketing Research*, **55**, 210–218.
- Train, K. (2003) Discrete Choice Methods with Simulation. *Cambridge University Press*.
- Yu, J., Goos, P., Vandebroek, M. (2009) Efficient conjoint choice designs in the presence of respondent heterogeneity. *Marketing Science*, **28**, 1, 122–135.
- Zwerina, K., Huber, J., Kuhfeld, W. F. (1996) A general method for constructing efficient choice designs. Durham, NC: Fuqua School of Business, Duke University.

Appendix A: the 20 choice sets used for the survey on television broadcasting

Sets	Month	Kickoff time	Opponent	Sets	Month	Kickoff time	Opponent
1	AUG	SUN 18:00	Top	11	AUG	SUN 18:00	Middle
	OCT	FRI 20:30	Top		JAN	SAT 20:00	Middle
	AUG	WED 20:30	Middle		OCT	SUN 14:30	Top
2	MAR	WED 20:30	Top	12	MAR	SAT 20:00	Lower
	JAN	SAT 18:00	Lower		JAN	FRI 20:30	Middle
	MAR	SAT 20:00	Middle		MAR	SUN 14:30	Middle
3	MAR	SUN 20:30	Lower	13	AUG	FRI 20:30	Middle
	AUG	SUN 18:00	Middle		JAN	SUN 18:00	Lower
	JAN	WED 20:30	Lower		AUG	SUN 14:30	Middle
4	MAR	SUN 14:30	Lower	14	JAN	FRI 20:30	Lower
	MAR	WED 20:30	Middle		AUG	SAT 20:00	Lower
	AUG	SUN 20:30	Middle		OCT	SUN 20:30	Middle
5	JAN	SAT 20:00	Middle	15	JAN	WED 20:30	Middle
	OCT	SAT 18:00	Lower		AUG	SAT 18:00	Top
	OCT	WED 20:30	Lower		JAN	FRI 20:30	Lower
6	JAN	SUN 14:30	Lower	16	OCT	SAT 18:00	Lower
	OCT	SUN 18:00	Lower		OCT	FRI 20:30	Middle
	JAN	SAT 20:00	Top		MAR	SUN 20:30	Top
7	AUG	WED 20:30	Top	17	OCT	SUN 20:30	Middle
	MAR	SAT 18:00	Middle		OCT	SUN 18:00	Top
	OCT	SAT 20:00	Top		MAR	FRI 20:30	Top
8	OCT	WED 20:30	Middle	18	AUG	SAT 20:00	Lower
	JAN	SUN 14:30	Middle		MAR	SUN 18:00	Middle
	AUG	FRI 20:30	Lower		AUG	SAT 18:00	Middle
9	AUG	SUN 20:30	Lower	19	JAN	WED 20:30	Top
	OCT	SUN 14:30	Middle		MAR	SAT 20:00	Middle
	JAN	SAT 18:00	Middle		OCT	FRI 20:30	Top
10	AUG	SUN 14:30	Lower	20	AUG	SUN 14:30	Top
	OCT	SAT 18:00	Top		MAR	SAT 18:00	Top
	OCT	SUN 18:00	Top		OCT	WED 20:30	Lower

Appendix B: the 16 choice sets used for the survey on stadium attendance

Sets	Month	Kickoff time	Opponent	Ticket type
1	OCT	SAT 20:00	Top	Roofed standing place
	JAN	SAT 18:00	Middle	Roofed standing place
2	AUG	WED 20:30	Middle	Non-roofed standing place
	OCT	SUN 14:30	Lower	Non-roofed seat place
3	OCT	FRI 20:30	Top	Roofed seat place
	JAN	SUN 14:30	Top	Roofed seat place
4	AUG	SAT 18:00	Top	Non-roofed standing place
	OCT	WED 20:30	Top	Roofed standing place
5	AUG	SUN 18:00	Middle	Non-roofed seat place
	MAR	SAT 20:00	Middle	Non-roofed seat place
6	JAN	SAT 20:00	Top	Non-roofed standing place
	MAR	FRI 20:30	Top	Roofed seat place
7	AUG	FRI 20:30	Lower	Non-roofed seat place
	OCT	SAT 18:00	Lower	Roofed standing place
8	OCT	SUN 14:30	Lower	Non-roofed standing place
	MAR	SUN 18:00	Lower	Non-roofed standing place
9	MAR	WED 20:30	Middle	Non-roofed seat place
	MAR	FRI 20:30	Lower	Roofed standing place
10	AUG	WED 20:30	Lower	Roofed standing place
	JAN	SUN 20:30	Middle	Roofed standing place
11	AUG	SUN 18:00	Middle	Non-roofed standing place
	OCT	WED 20:30	Lower	Non-roofed standing place
12	AUG	SAT 20:00	Top	Non-roofed seat place
	AUG	SUN 14:30	Top	Non-roofed standing place
13	AUG	SUN 20:30	Top	Roofed standing place
	OCT	FRI 20:30	Middle	Non-roofed seat place
14	AUG	SUN 20:30	Top	Non-roofed seat place
	JAN	SAT 20:00	Top	Roofed seat place
15	AUG	FRI 20:30	Middle	Roofed standing place
	JAN	SUN 18:00	Lower	Roofed seat place
16	AUG	SAT 20:00	Lower	Roofed seat place
	JAN	FRI 20:30	Top	Roofed seat place