Innovative applications of O.R

Armchair fans: Modelling audience size for televised football matches

Babatunde Buraimo[a], David Forrest[a,*], Ian G. McHale[a], J. D. Tena[a,b]

[a] Centre for Sports Business, University of Liverpool Management School, Liverpool, L69 7ZH, United Kingdom
[b] DiSea & CRENOS, Università di Sassari, 07100, Sassari, Italy

ARTICLE INFO

ABSTRACT

Modelling audience size for televised football matches has utility for a number of stakeholders. In the short-term it can inform the choice of broadcasters as to which matches to show and decisions of firms regarding how much to bid for advertising slots during the game. For elite clubs and leagues, for which television rights is their principal source of revenue, it offers a means of understanding which match characteristics will draw the largest audience and this should inform long-term decisions on how to structure their competitions. The paper argues that previous modelling in this area has typically depended on an inappropriate measure of audience size and unsatisfactory metrics to capture important match characteristics such as the quality of the talent on show and the significance of the match for the championship or relegation. In contrast to this prior literature, the present paper draws on sport analytics to derive its measures of player talent and match significance. This makes a material difference to findings from modelling audience size for 790 English Premier League matches between 2013 and 2019. For example, it identifies elevated interest for matches potentially significant for end-of-season outcomes. This has implications when organisers consider the optimal structure of the competition and revenue sharing arrangements between member clubs of the league. For the wider television industry, we show that measurement of audience size (and some results from modelling) are sensitive to whether focus is on the whole programme or just its core content (in this case the match itself).

© 2021 Elsevier B.V. All rights reserved.

1. Introduction

The notion of applying operational research to problems in sport originated in the 1950s when it was argued that certain sports could be compared structurally with military campaigns and should therefore be amenable to analysis which would help improve strategy and tactics (Wright, 2009). Thus, early applications were mainly for use by coaches (Coleman, 2012). Here, we present analysis and modelling that has consequences for a much wider range of stakeholders and will inform decision making in the business of sport, sports broadcasting and marketing, and even help those charged with structuring sport leagues. The paper considers the task of understanding what attracts football fans to watch a match on television.

Studies in the literature on modelling the audience for sport have obvious utility to various stakeholders. For example, they may be useful to broadcasters when decisions are taken on which matches to select for showing on television and to advertisers interested in the likely audience size for upcoming televised games, relevant, for example, to bid pricing decisions for advertising slots either during the programme or at the stadium.

Because results from these studies reveal which match characteristics enhance viewing figures, findings may also be used to inform sports leagues' decisions on how best to design the formats of their competitions to optimise audience size. In elite sports leagues, television audiences represent the most lucrative source of revenue. For example, in season 2018–19, the EPL's revenue from broadcasting rights was almost exactly four times that from 'match day income' (Deloitte, 2020, p. 16). A similar business model applies beyond football, in most American Major League Sport and in cricket's Indian Premier League.

The first academic journal article modelling the size of television audiences for sport (Forrest, Simmons & Buraimo, 2005) focused on the English Premier League (EPL). Since then there has been a large successor literature on the same theme, embracing both football and many other sports. Van Reeth (2020) provides a useful tabulation of studies to date. A high proportion focus particularly on the 'outcome uncertainty hypothesis', which, at the match level, proposes that interest will be enhanced where there is greater ex ante uncertainty over which team will win. Indeed, previous studies of television demand in the context of the EPL have

included 'uncertainty' in their title (Buraimo & Simmons, 2015; Cox, 2018; Forrest et al., 2005), signalling their primary motivation. The same emphasis is found in the paper titles of television audience studies for other football leagues (e.g. Pérez, Puente, & Rodríguez, 2017; Schreyer, Schmidt & Torgler, 2016) and for other sports as well, for example Formula 1 motor racing (Schreyer & Torgler, 2018), cycling (Van Reeth, 2019) and Australian Rules Football (Dang, Booth, Brooks & Schnytzer, 2015). That it is a recurring theme may reflect that sports leagues have argued that consumers' interest in closely contested contests justifies measures which would otherwise be inconsistent with competition law, for example revenue sharing and salary caps. Meanwhile, Scarf, Parma and McHale (2019) showed that rules designed to heighten entertainment by increasing scoring rates in rugby union, have a detrimental effect on outcome uncertainty. Amongst other insights, modelling audience size provides a means to testing whether the emphasis on outcome uncertainty is justified.

Our reason for offering a new study of audience size is that, with the benefit of hindsight, we believe that previous modelling has been flawed and that conclusions drawn from it may therefore be spurious. In particular, studies appear often to choose an inappropriate measure for the dependent variable, audience size, and inadequate proxies for player quality and match significance. We will argue that previously adopted metrics fail meaningfully to capture the underlying concepts of interest and their use may lead to bias in the estimation of key variables, including outcome uncertainty. In proposing alternative metrics, we will draw on advances in sport analytics and demonstrate that there are material differences in findings when measures derived from sport analytics are employed in preference to ad hoc measures used in previous research.

Section 2 will identify problems in the prior literature including issues around measurement of audience size, player quality and match significance. In Section 3, we propose and explain our alternative metrics. Section 4 describes the data available to us for analysis and builds the statistical model to be employed in investigating determinants of audience size. Findings are presented and discussed in Section 5 and Section 6 offers closing remarks on the significance of our results.

2. Problems in prior literature

2.1. Empirical framework

Studies of the size of the television audience for football at the level of the match have generally been framed in terms of a model such as:

$$\text{Ln(audience size)} = f(\text{player quality, outcome uncertainty,}$$
$$\text{match significance, controls}) \quad (1)$$

In the first paper to model audience size, Forrest et al. (2005) measured outcome uncertainty by the current difference in points per game between the opposing clubs, adjusted to take account of the average points value of home advantage in the league. They reported this measure as significant (though with small effect size). Nevertheless it may be regarded as a fairly crude measure since it fails to take into account factors influencing prospects for the match which are known to the prospective audience but are not captured by summary statistics of the performances of the two clubs over the whole of the season to date. For example, a club may have strengthened its team by entering the transfer market in the January window precisely because it was dissatisfied with its current position in the standings and its probability of a win in February may then be much greater than its current points total would suggest.

Most recent studies have therefore measured uncertainty using betting odds, which are presumed to capture all relevant current information. Accordingly, the most common measure of outcome uncertainty adopted in recent studies (e.g. Bergmann & Schreyer, 2019; Buraimo & Simmons, 2015; Caruso, Addesa & Di Domizio, 2019; Cox, 2018) is the absolute difference in the win-probabilities of the home and away clubs, as implied by the betting market. In this paper we do not challenge this consensus and will adopt the same metric in our modelling. However, we identify problems in all the other elements of (1) and will now address each in turn.

2.2. Measuring the dependent variable

There is nothing to which to object in a decision to adopt the natural log of audience size as the dependent variable. However, this begs the question (even if few authors answer it): what is audience size? Van Reeth (2020) draws attention to how ambiguous this term can be. In Europe at least, audience research agencies typically compile estimates for the number of viewers at each minute of each programme of each television channel, based on information transmitted from 'peoplemeters' installed in a representative panel of households. Estimates of audience size for a given programme can then refer to the average per-minute audience or to the peak audience or to either expressed as a percentage of all viewing of programmes being broadcast at the time. Nearly all published studies on football focus on average per-minute-audience (though Schreyer et al., 2016, also present results for share of all viewers watching television at the time). But this still leaves ambiguity. The headline figure for audience size from the audience research agency will relate to the whole programme whereas the match is only a sub-set of the programme. Almost always, there will also be pre- and post-match content (build-up before kick-off, analysis and interviews after the game).

We have ascertained that several studies, including Forrest et al. (2005), used the average audience across the programme rather than during the match as the dependent variable. Some other published papers are not explicit in defining their dependent variable but we suspect that most (Pérez et al., 2017 is an exception) have used a programme-based measure because anything else is expensive to procure. This is a problem because the duration of the pre- and post-match segments may vary substantially across matches. For example, in the data set we employ here, the mean duration of the post-match segment for weekend matches was 25 min; but the standard deviation was 21 min and, after one game, the programme continued for 79 min beyond the final whistle. Audience size during programmes is known to be appreciably higher while the match is in progress and so including pre- and post-match content in the measurement will tend to shrink the estimate of the average audience for the programme by a greater proportionate amount where the programme has been longer. This will introduce 'noise' into the data, making estimation less precise. Worse, it may lead to biased coefficient estimates if the length of the programme is correlated with covariates. Such correlation is plausible, for example broadcasters may schedule longer coverage for a fixture with high 'match significance' or for a game featuring popular clubs. This is the first problem which we shall seek to correct in our empirical analysis.

2.3. Measuring player quality

Papers in the literature, while often focusing on the importance (or not) of outcome uncertainty, invariably recognise that the quality of play is likely to matter and include in their models a measure intended to capture the level of talent featuring

in the particular match. The proxy proposed by Forrest et al. (2005) was the combined wage bill of the two clubs for the season in question (relative to the average for all clubs that season). This 'standardised wages' metric has continued to feature in more recent studies (e.g., Buraimo & Simmons, 2015; Caruso et al., 2019; Scelles, 2017).

This use of club wage bills is convenient because they are available from clubs' financial accounts and may have loose rationale from its underlying assumption that labour markets for talent will be s efficient. However, a number of concerns may be raised. First, published wage bills include remuneration for non-playing staff and cover different squad sizes: unused players increase the wage bill but do not appear on the field and so do not raise the amount of talent to be viewed by the television audience. Second, published wage bills include substantial bonus payments, linked to personal or team performance or number of appearances as calculated at the end of the season. Third, some players may be unavailable to play in a particular match, e.g. because of injury or suspension Fourth, individual wages were set at the time each player signed his contract, typically up to five years before, and so his current wage may not accurately reflect his current ability. Fifth, some transfers-in might have proved misjudged; and some players may have been signed with the club already anticipating a downward age-related trajectory of performance (but with wages smoothed out across the contract). Sixth, each club's wage bill is invariant through a season and unresponsive to fresh information Indeed, a given annual wage bill may conceal very different wage bills through the season if clubs have bought or sold high value players in the transfer window. Seventh, wage bills for larger market clubs will be expected to be persistently higher than for small market clubs. Use of club fixed effects, to reflect the power of football brands, will then be problematic because of high correlation with wage bills.

Other studies have employed alternative proxies for quality. For example, Schreyer et al. (2016) used the sum of the 'transfer values' of the two starting elevens in the match. Although not explicitly stated in their paper, we take these as (crowd-sourced) values from the transfermarkt website. The advantage over club wage bills is that this metric can be based solely on valuations of players taking part in the particular fixture. However, Coates and Parshakov (2021) found that transfermarkt values were biased and failed to reflect playing performance, as measured by simple metrics such as goals scored and assists.

Yet other researchers have attempted to capture quality through a performance metric for the club rather than sum across players, for example, Cox (2018) uses goals scored and conceded in the last six fixtures. This measure may be distorted by short-term loss of team form and has the further disadvantage of failing to account for differences in team composition, and of the strength of opposing teams, between recent fixtures and the present fixture.

We will argue that advances in sport analytics now enable player talent in a match to be measured directly, making it unnecessary to rely on imperfect money proxies or ad hoc summary statistics of past team performances. Our measure of quality will sums ratings of individual players actually taking part, ratings obtained directly from their past on-field contributions and updated at every round of matches. We would expect viewing decisions to be better explained by up-to-date information captured in appropriate performance metrics. This seems a reasonable assumption since a sizeable proportion of the potential audience follows the sport closely: viewing requires payment of a subscription to at least one broadcasting service, which is likely to limit the number of viewers with only marginal interest. We will demonstrate that audience demand shows great sensitivity to our metric and its employment makes a substantive difference to estimation results.

2.4. Measuring match significance

Match significance refers to the importance of the fixture to seasonal outcomes. As in other European top-tier divisions, the EPL offers three levels of 'prizes' awarded according to league positions at the end of the competition. The club with the greatest number of points becomes Champion. It and the next three clubs are awarded places in the European Champions League in the following season. The bottom three of the twenty clubs receive a negative prize, relegation into the second-tier league, with considerable loss of prestige and huge loss of revenue.

It is a reasonable hypothesis, to be tested, that audience interest is stimulated when a particular fixture is expected to have a strong relevance to one or more of these seasonal outcomes both because viewers are themselves interested in seasonal outcomes and because they may anticipate that important matches will be fought more intensely. This justifies the inclusion in some previous studies of variables to represent match significance for the championship, European qualification and relegation. Unfortunately, attempts to construct such variables for use in television audience (or stadium attendance) studies have not yielded measures which can be regarded as credible.

Attempts to produce viable proxies for match significance in football began with Jennett (1984) who included championship significance and relegation significance in an attendance model for Scottish games. Each was compiled in a different way but each was open to the criticism that it required information available only ex post by reference to points totals in the league table at the end of the particular season.

Later authors have suggested ad hoc measures which require only information available at the time of the match. For example, adapting an algorithm proposed by Goddard and Asimakopoulos (2004), Buraimo and Simmons (2015) employed dummy variables signalling a match as significant for the championship for a particular club if it were still possible for it to win the title if it secured victories in all its remaining fixtures while all other clubs averaged just one point per game. Forrest et al. (2005) included dummies referencing matches where one or both clubs were in positions in the league table around the regions where prizes would be awarded at the end of the season. However, the use of these ad hoc variables failed to reveal a significant role for match significance in accounting for size of television audiences.

A potential explanation may be that the dummy variables employed were insufficiently discriminating. For example, early in the season, all clubs are technically 'in contention' for all the prizes on offer from the League even though, for example, the prospect of winning the Championship may be utterly unrealistic for many of them. This leads to too many matches being categorised as 'significant' (as in Buraimo & Simmons, 2015) or else applying the dummy variable only to matches later in the season (as in Forrest et al., 2005), which will miss earlier games between, for example, the clubs most favoured for the Championship. Again, the dummy variables are insensitive to how many matches remain to be played by a supposed contender for a prize and to which opponents are yet to be faced. To illustrate, a club may technically still be in contention for the Championship if it would take the prize by winning its remaining seven matches while those ahead of it took only one-point-per-match but its probability of winning the League would nevertheless be very low and unlikely to change much from its upcoming fixture.

Faced with these problems, some recent studies-such as Cox (2018) on the EPL and German and Italian studies by Schreyer et al. (2016) and Caruso et al. (2019) respectively- omit match significance altogether. But either using a weak metric for match significance or leaving it out is unsatisfactory. First, the chance is lost to answer important questions for leagues, such as whether they

should restructure themselves to exploit a demand for significant matches. Second, coefficient estimates on other important variables may become biased because of their correlation with (true) match significance. For example, the most significant matches will probably be between clubs adjacent to each other in the table (at the top or bottom) and this will be reflected in the value of the outcome uncertainty measure.

Prior studies employed metrics which can be arithmetically derived from league tables but failed adequately to capture match significance. We contend that measurement is better approached through sport analytics. This approach has not been attempted before in modelling television audiences for football. However, a baseball stadium attendance paper (Tainsky & Winfree, 2010) included a variable to measure the importance of a match for play-off prospects. Their metric, applied only in the second half of the season, used a forecasting model based on current team win-percentages to simulate the rest of the season from the time of the subject match. Below we will employ an indicator derived from a forecasting model and simulation of the rest of the EPL season to infer how much the current match matters for the championship, European qualification and relegation.

3. New measures for player quality and match significance

3.1. Player quality

Rating players in football is most often done by attributing values to the actions the players perform on the pitch. For example, McHale, Scarf and Folker (2012) describe how actions such as passes and tackles can be valued by estimating the relationship these actions have with generating shots. Nowadays, rich data on the timing and location of events on the field allow actions to be valued in even more detail. For example, Liu, Luo, Schulte and Kharrat (2020) use reinforcement learning to value each action in a match by estimating its influence on a team's chances of the current possession ending in a goal. But these methods require huge amounts of data and are computationally expensive. For a single season for just one league, there are millions of data points meaning that not everyone can adopt such ratings models despite their quality.

Instead of valuing actions to rate players, here we adopt an approach presented in Kharrat, McHale and Peña (2020). They adapt the concept of plus-minus ratings for use in football. Plus-minus ratings have been used in basketball and ice-hockey for more than fifty years. The basic premise is simple and intuitive: a player's plus-minus rating quantifies his effect on his team's performances by comparing the team's performances when the player is on the pitch to when the player is not. This idea has several associated complications: first, one must decide on how team performance is measured; and second, one should control for the strength of the other players on the pitch (both on the player's own team and the opposition side).

To address the issue of accounting for the strength of the other players on the pitch, Rosenbaum (2004) estimated plus-minus ratings using a regression framework. In this framework, the dependent variable is a measure of the team's performance (e.g. in basketball, point differential per minute is used) and each observation is a segment of play in which the set of players on the pitch is constant. A new observation 'begins' with each substitution, each red card, and each new match. The covariates are dummy variables for the identity of all players in the league. For a given segment of play, a value of 1 indicates the player is on the home team, a value of −1 indicates the player is on the away team, and a value of 0 is given for players not on the pitch. In doing this, one can simultaneously account for the quality of the player's teammates and the opposition players, and also home advantage.

Although Rosenbaum's (2004) solution worked well for basketball, team line-ups in football change infrequently, and by very little during an individual match. This means that covariates are highly correlated. To solve this problem in the context of ice hockey, MacDonald (2012) used ridge regression to estimate the parameters of the model.

The task of measuring team performance may, at first, seem simple: in basketball, use the point difference between the two teams during the segment of play, and in soccer, use the goal difference between the two teams in a segment of play. However, unlike basketball, football is a low scoring game. This means that using goal difference as the dependent variable results in a sparse response variable that is mostly 0. To solve this problem, Kharrat et al. (2020) proposed two alternative 'performance metrics' which are less sparse. Here we use their 'expected points' plus-minus ratings model, where 'points' refers to the 3 points which are awarded to the team winning the match, 1 point to each team in a drawn match, and 0 points to the team which loses the match. The idea is that, during a segment of play in a match, the points the team is expected to win changes with both the passing of time, and the occurrence of goals. It is this change in expected points during the segment that is used as the dependent variable. If a player is substituted on to the pitch when the scoreline is 5–0, and the match finishes 6–0, the change in expected points is very small since at 5–0, the team was already almost certain of winning the match. On the other hand, if a player is substituted on to the pitch when his team is 0–1 down, and the game finishes with a 2–1 win for his team, then he was playing when there was a considerable positive change in the expected points. The resulting ratings identify players who improve their team's results, and the authors present tables of the top players in each season across European football leagues. The ranking of top players would not be surprising to followers of the sport, so the method appears to have face validity. Indeed, Premier League clubs have started to use the ratings to help in recruitment decisions on players.

We use expected points plus-minus ratings to represent the quality of the players on each team. The ratings are calculated using matches in the 12 months prior to the game such that the ratings are truly out-of-sample. The quality on display is represented by the average expected points plus-minus rating of the starting 22 players of the match.

3.2. Match significance

Conceptually, following Schilling (1994) and similar to Scarf and Shi (2008), we evaluate the significance of an upcoming match for a particular club in terms of how much difference it would make to its probability of securing a seasonal prize if the fixture were won rather than lost. We then define the overall significance of a given fixture by adding together the significance measure for each club, performing separate calculations for each of the prizes to be awarded: (i) the championship, (ii) European qualification (top four clubs) and (iii) relegation (bottom three clubs).

Formally, for match i, the match significance for the outcome of the championship is

$$matchSig_i^{championship}$$
$$= \Pr\left(home\,team\,wins\,the\,championship | home\,team\,wins\right)$$
$$- \Pr\left(home\,team\,wins\,the\,championship | home\,team\,loses\right)$$
$$+ \Pr\left(away\,team\,wins\,the\,championship | away\,team\,wins\right)$$
$$- \Pr\left(away\,team\,wins\,the\,championship | away\,team\,loses\right).$$

(2)

Similar calculations were made for both the identity of the top four clubs and the identity of the bottom three. In the case of

a match relevant for relegation, a win rather than a loss would lower the probability of a club finishing in the bottom three, making the significance measure negative. For 'relegation significance', we therefore take the absolute value of expression (2) as our metric.

To calculate the probabilities, we simulate the results of all remaining games in the season 100,000 times. For this, we need to employ a match forecasting model. We use the model first proposed by Maher (1982), which assumes the scoring rates (the number of goals they score in a match) of the two teams follow Poisson distributions such that

$$X \sim Poisson(\gamma \alpha_H \beta_A) \; and \; Y \sim Poisson(\alpha_A \beta_H) \tag{3}$$

where α_H is the attack strength of the home team, and β_A is the defence strength of the away team. γ is a parameter allowing for home advantage.

Rather than estimate the values of the parameters using data on the scorelines of past matches, we backward engineer the values of the parameters from the bookmakers' odds on the outcomes of the matches. Specifically, the odds for home win (H), draw (D), away win (A), over 2.5 goals (O), and under 2.5 goals (U) are used. For each match we had odds on five markets. We use average bookmaker odds for each market, collected by football-data.co.uk. We minimise the following 'error' function which reflects the squared distance between the probabilities implied by the bookmakers and the probabilities implied by the double Poisson model

$$f(\theta) = \sum_{k \in M_t} \left(\sum_{i \in H,D,A,O,U} (bookie_{ik} - model_{ik})^2 \right)^{\exp(-\zeta(t-t_k))} \tag{4}$$

where M_t is the set of all matches in the season up to time t, t_k is the date the kth match is played. $bookie_{ik}$ is the bookmaker implied probabilities for the ith market (home win, draw, away win, over 2.5 goals, under 2.5 goals) for the kth match. As the season progresses, bookmakers re-evaluate their estimates of the relative strengths of the teams. To account for more recent matches reflecting the current assessment of teams' abilities, we follow Dixon and Coles (1997) and include a time decay factor in the error function. The specification means that the half-life of the decay factor is $\ln(0.5)/-\zeta$. Following Dixon and Coles, we use $\zeta = 0.002$ such that the half-life is around 350 days.

The advantage of using backward engineering is that team strengths can be estimated quickly, and accurately, early on in the season. We estimate the parameters for all teams from 1st September of each season (strictly speaking, the parameters could be estimated once each team has played at least one match; however, the estimated strengths will be unconnected; after each team has played two games, links exist so that the parameters can be estimated and be relative to all other teams). Typically around 2 to 3 rounds of matches have been played at this point.

$f(\theta)$ is minimised with respect to the 40 parameters of the model (one home advantage parameter, 19 attack parameters and 20 defence parameters; one parameter must be kept constant to avoid over-parameterisation and we choose Arsenal's attack parameter). The parameters are re-estimated after each round of games, and used to simulate the remaining matches in the season.

The 2018–19 Premier League season saw a titanic struggle between Liverpool and the eventual champion, Manchester City. Both clubs were able to win the championship up to the very last game of the season. To illustrate the data generated, Fig. 1 shows the probability of each team winning the league title if it wins the next game, and if it loses the next game. Towards the end of the season, the result of the next game has a large impact on

the outcome of the championship. Indeed, Liverpool had to win the final game to even have a chance of winning the title. Fig. 2 shows the resulting values of match significance for each team separately as the season progresses. The match between the two on 3rd January 2019 is visible as a local peak in match significance for each of the clubs (and therefore for our match metric which sums the figure for each club), demonstrating its importance to the championship several months before the end of the season.

4. Data and model

4.1. Data

We had access to minute-by-minute estimates of the size of the domestic audience for every programme featuring a live EPL match between the middle of the 2013–14 season and the end of the 2018–19 season. These were sourced from the British Audience Research Bureau (BARB), which supplies the broadcasting and marketing industries with data on television viewing in the United Kingdom. Its audience figures are based on a representative panel of more than 5,000 households covering 12,000 individuals over the age of 4. As it is a panel of households, viewing in other settings, such as bars or prisons, is not reflected in BARB's figures for audience size.

During the period, matches were shown on either of two channels, Sky Sports and BT Sport, each accessible to viewers by payment of a subscription (Sky subscribers can view BT matches by paying a supplementary fee). Sky Sports has been supplying EPL matches to its customers since the inception of the EPL but BT Sport was in its first season as an EPL broadcaster at the start point in our data. That season, and in each of the following two, 154 EPL matches were televised on one or other of the channels. This degree of exposure was increased only modestly for the final three seasons, when 168 matches were shown.

For estimation, we discarded matches played in the first three rounds of each season and in the last round of each season. Information from all matches in the first three weeks was used to calibrate our match significance variables. In the last round of each season, to assure the integrity of the competition, all matches are played simultaneously and more than one game is shown on television. This makes conditions different from the rest of the season when any match to be televised is scheduled to a time slot such that no other EPL fixture is being played at the same time (the most common time slots are lunchtime and late afternoon on Saturday and Sunday and starting around 8 p.m. on Monday). These omissions from the data left us with 790 televised matches to be included in regression analysis.

For each match, we had minute-by-minute audience size data for the programme. As noted above, programme length is highly variable and therefore our preference was to model average per-minute audience for the match itself, which is of fixed duration, rather than for the programme of which it is the centre-piece. Our preferred dependent variable is therefore the natural log of the average per-minute audience size measured from kick-off to 110 min later. This time interval accounts for 90 min of regular play, 15 min interval (half-time) and an assumed 5 min of added (injury) time. For comparison with earlier studies, we also modelled programme audience size. This is the average per-minute audience size for the whole programme, which includes pre- and post-match content as well as the game itself. This is the 'headline' statistic which will usually be quoted in the media. Its mean across our data (829,862) was appreciably lower than the mean audience for the match itself (1.06 millions), reflecting that pre- and post-match segments typically attract far fewer viewers than the period of action on the field.
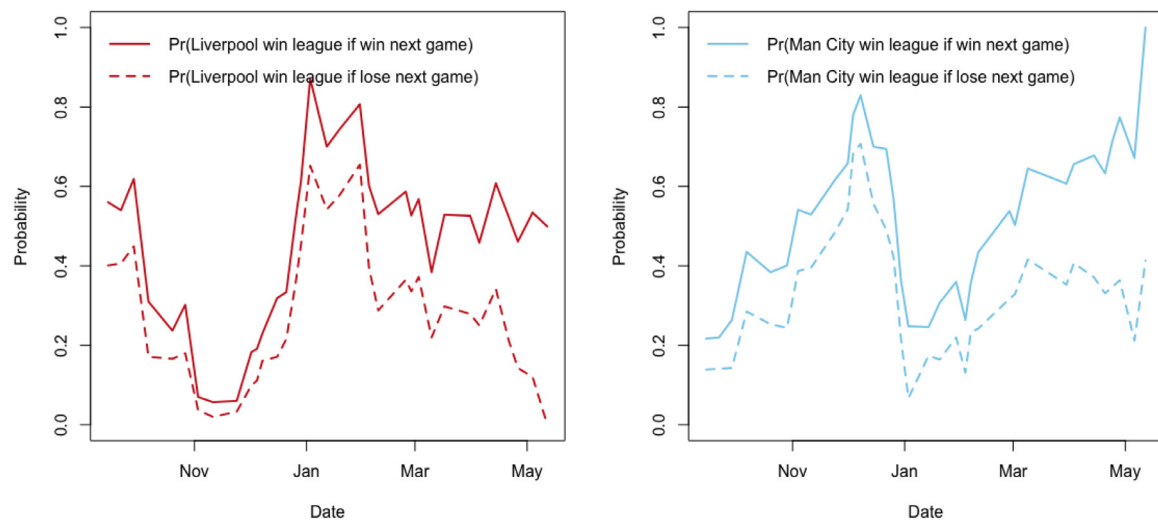
Fig. 1. Probabilities of winning the championship conditional on the next game being won (solid line), and conditional on the next game being lost (dashed line), for Liverpool (left), and Manchester City (right), during the 2018–19 Premier League season.
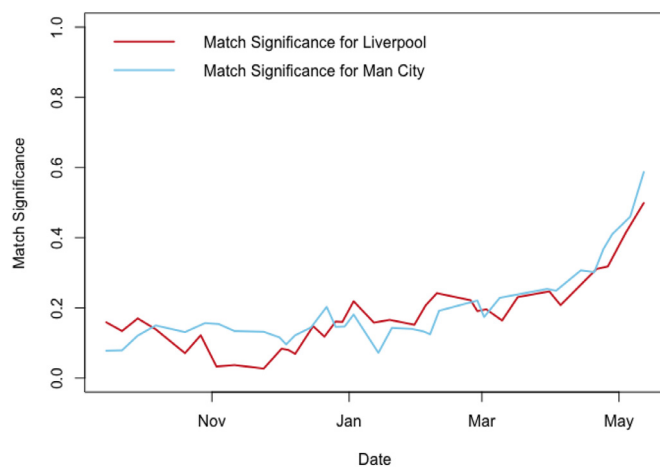

Fig. 2. Match significance for Liverpool (red) and Manchester City (sky blue) during the 2018–19 Premier League season.

4.2. Model

The regressors of key interest in our model are player quality, outcome uncertainty, and three variables for match significance (representing the relevance of a match for the championship, European qualification and relegation).

Outcome uncertainty is measured as the absolute difference in the probabilities of a win for either team, according to bookmaker odds. We retrieved odds offered by William Hill as displayed in the archive at football-data.co.uk. These were expressed in decimal-odds format and the inverse of the decimal-odds gave us the 'bookmaker-probability' of each outcome (home win, draw, away win). Finally, since the sum of the three bookmaker-probabilities always exceeds 1, to allow the betting provider its commission, it was necessary to rescale the bookmaker-probabilities such that they added to 1, which we did by multiplying each by a constant. This method for obtaining 'implied probabilities' was termed 'basic normalisation' by Štrumbelj (2014).

The variables representing player quality and match significance were described in Section 3 above. The expected signs on outcome uncertainty, player quality and the three match significance variables are all positive.

We also include several control variables. These include dummy variables to represent the time of the week and the month of the year when the game was played, the season during which it took place and the broadcaster which provided the coverage. There is also an indicator for 'derby match'.

Regarding time in the week, we distinguish weekday matches (always in the evening) from weekend matches (the reference category), in line with previous studies for the EPL. In addition, we distinguish a third category, 'Christmas', which refers to the period following Christmas Day and up to the day of the New Year Holiday (as late as January 3 in 2017, because the 1st had fallen on a Saturday). All matches in this period, when a large part of the labour force is on holiday, are deemed 'Christmas' rather than counted as weekday or weekend. The delineation of this third category is made because, controversially, British leagues schedule frequent fixtures over this time rather than take a midwinter break as in most of the rest of Europe. The alternative would be to schedule extra rounds of midweek matches during the rest of the season and it is relevant to ask whether there is any gain if the goal is to maximise aggregate television audience.

Audience size is likely to depend on which broadcaster shows the match. 'BT Sport' is a dummy variable to distinguish its games from those covered by Sky Sports. We also include interaction terms between BT Sport and season dummies because BT Sport was a new entrant to the market in the first season of our data period and it would be reasonable to suppose that its penetration of the market would be spread over time. Coefficient estimates on the interaction terms would also reflect any differential price changes compared with the long-time incumbent, Sky Sports. We were unable to track these though we do note that, on a per-match basis, BT Sport subscription prices were much higher than Sky Sports, 2.6 times as high in season 2015–16 according to Butler and Massey (2019). In a private communication, Dr. Butler informed us that the differential was similar in the most recent season.

'Derby match' also features in most earlier studies. This indicator variable signals match-ups between clubs where there is local or regional rivalry. We identified 16 such match-ups, eight of which involved London clubs. Of the remainder, most related to clubs in contiguous urban areas but we also included three match-ups (Liverpool-Manchester United, Brighton-Crystal Palace and Brighton-Bournemouth) where there was greater geographical separation but where common knowledge still recognises strong rivalry. We expected that derby matches might attract additional

viewers because of extra regional interest in the relevant matches but possibly also on a wider geographical basis because of the perception that these games are contested more intensely.

Finally, and similar to, for example, Buraimo and Simmons (2015), we include a full set of club dummies, each set equal to 1 if the relevant club was a participant in the subject match. Some other authors, for example Pérez et al. (2017) and Forrest et al. (2005) were more selective in that they each included dummy variables only for two or three clubs with national reach in support (in Spain and England respectively). However, clubs with very different market sizes played in the EPL over our period, including some which appeared to maintain a historically strong support base but which were not now strong either financially or on the field. Sunderland is an example. Representing a club like Sunderland by its own dummy variable allows us to estimate the power of club brands to draw audiences independent of the quality of their current players. The reference club, selected on lexicographic grounds, is AFC Bournemouth, one of the smallest market clubs in the EPL and in fact only a recent entrant to the EPL (historically, it had most often played in the third-tier league). Because Bournemouth would be the point of reference, we anticipated that most club dummies would attract a positive coefficient estimate.

Our model to be estimated is:

$$\text{Ln}\,(audience\,size) = f\,(\text{player quality, outcome uncertainty,}$$
$$\text{match significance (championship),}$$
$$\text{match significance (European),}$$
$$\text{match significance (relegation), controls)} \quad (5)$$

where controls include: club dummies, season dummies, month dummies, derby match, BT Sport, BT Sport/season interaction terms.

Because not all EPL matches are screened on television, we considered the case for trying to account for possible selection bias when estimating this model. Two previous papers (Buraimo & Simmons, 2015; Forrest et al., 2005) modelled broadcaster choice of which matches to show, incorporated into a Heckman procedure. Others have ignored the issue (it did not arise in studies for Germany and Italy, where all matches were televised). We decided not to employ a sample selection model here, for three reasons. First, there is no strong theoretical basis for suspecting that selected matches possess some distinctive non-observed characteristics which would affect audience size given that the set of match characteristics already included in the regression equation seems to be rather comprehensive. Second, preceding papers which have tested for sample selection bias have decisively rejected its presence. Third, we are not confident that we would represent the process driving broadcaster choice accurately because they will have selected their games at unknown dates several weeks before each round of matches takes place (to allow fixtures to be rescheduled to the time slots reserved for televised matches), so the dating of relevant covariates would be problematic. Another complication of modelling broadcasters' choice of matches to be screened would be that the contracts require them to choose four from ten predetermined fixtures in each round. Preceding literature fails to account for this constraint and treats the choice of matches as if it were made from all games in the season.

Tables 1a and 1b present summary statistics for variables included in the modelling. In the data set, the match with the highest average audience (measured over the match rather than the programme), 2.67 millions, was Chelsea v. Manchester United, played in January 2014.

Table 1a
Summary statistics for continuous variables, $N = 790$.

	mean	std. dev.	min	max
programme audience	829,682	380,248	171,500	2432,500
match audience	1062,880	468,241	181,199	2672,705
average player rating	0.013	0.012	−0.020	0.047
combined relative wages	2.410	0.767	0.872	4.470
outcome uncertainty	0.341	0.217	0.000	0.885
match significance (championship)	0.063	0.101	0.000	0.995
match significance (European)	0.136	0.133	0.000	0.920
match significance (relegation)	0.088	0.125	0.000	1.597

Table 1b
Summary Statistics for discrete variables, $N = 790$.

	mean	std. dev.
derby match	0.117	0.322
Christmas	0.060	0.237
Weekday	0.213	0.409
October	0.095	0.293
November	0.094	0.292
December	0.148	0.355
January	0.111	0.315
February	0.108	0.310
March	0.108	0.310
April	0.163	0.370
May	0.086	0.281
BT	0.257	0.437
BT × season ending: 2015	0.044	0.206
BT × season ending: 2016	0.043	0.203
BT × season ending: 2017	0.049	0.217
BT × season ending: 2018	0.049	0.217
BT × season ending: 2019	0.048	0.214
season ending: 2015	0.172	0.378
season ending: 2016	0.171	0.377
season ending: 2017	0.189	0.391
season ending: 2018	0.190	0.392
season ending: 2019	0.187	0.390

5. Findings

5.1. Match audience versus programme audience

Table 2 displays results from three models, each estimated with club dummies. Column (1) represents our preferred model, with the dependent variable measuring average audience size over the match period. Model 2 presents results with average audience size defined over the programme rather than the match itself. Model 3 replicates our preferred model except that the quality of players in the match is proxied by the clubs' total wage bill, an alternative measure popular in earlier papers. Covariates are not standardised but extensive reference to effect sizes will be made in the commentary. For our preferred Model 1, Fig. 3 plots observed and fitted values.

It was noted earlier that results in preceding studies appear typically to have measured average audience size over the whole programme rather than whistle-to-whistle and that this was a risky procedure because duration of programme is highly variable. Table 2, column (2) presents estimates based on average programme rather than average match audience size. Measuring audience size over the match itself results in more precise estimation. The standard deviation of the match audience variable (468,241) is larger than the standard deviation of the programme audience variable (380,248) but still the root mean square error of the match audience equation is appreciably lower. This encourages us to believe that modelling based on the audience just for the match itself should allow more reliable inference concerning viewer preferences. Differences in substantive findings on focus variables are limited; but modelling match audience sharpens coefficient esti-

Table 2
Regression results. dependent variable: Ln(audience).

	(1) match audience		(2) programme audience		(3) match audience	
	coef.	|t|	coef.	|t|	coef.	|t|
average player rating	3.971***	(2.92)	4.056***	(2.94)		
combined relative wages					0.040	(0.71)
outcome uncertainty	−0.037	(0.79)	−0.011	(0.22)	−0.031	(0.67)
derby match	0.051*	(1.89)	0.050	(1.60)	0.052*	(1.88)
match significance (championship)	0.675***	(5.21)	0.576***	(4.49)	0.788***	(6.18)
match significance (European)	0.206*	(1.95)	0.173*	(1.71)	0.226**	(2.13)
match significance (relegation)	0.341***	(3.90)	0.344***	(3.40)	0.274***	(3.06)
Christmas	0.100***	(2.87)	0.146***	(3.70)	0.093***	(2.71)
Weekday	0.016	(0.81)	−0.129***	(6.34)	0.015	(0.76)
October	0.017	(0.46)	0.027	(0.64)	0.018	(0.47)
November	0.106***	(3.06)	0.128***	(3.22)	0.105***	(3.01)
December	0.121***	(3.58)	0.146***	(3.78)	0.120***	(3.54)
January	0.195***	(5.63)	0.228***	(6.06)	0.196***	(5.57)
February	0.119***	(3.59)	0.161***	(4.38)	0.117***	(3.54)
March	0.052	(1.40)	0.102**	(2.47)	0.053	(1.37)
April	0.047	(1.47)	0.094***	(2.70)	0.043	(1.31)
May	−0.141***	(2.67)	−0.055	(1.10)	−0.149***	(2.83)
BT	−0.781***	(16.14)	−0.859***	(14.94)	−0.780***	(15.99)
BT × season ending: 2015	0.131**	(2.22)	0.207***	(2.95)	0.125**	(2.10)
BT × season ending: 2016	0.253***	(4.08)	0.283***	(3.94)	0.239***	(3.76)
BT × season ending: 2017	0.305***	(4.94)	0.354***	(5.07)	0.300***	(4.84)
BT × season ending: 2018	0.378***	(5.87)	0.427***	(5.85)	0.375***	(5.76)
BT × season ending: 2019	0.278***	(4.38)	0.351***	(4.95)	0.281***	(4.40)
season ending: 2015	−0.090**	(2.30)	−0.138***	(3.10)	−0.084**	(2.19)
season ending: 2016	−0.170***	(4.09)	−0.176***	(3.98)	−0.170***	(4.19)
season ending: 2017	−0.295***	(7.51)	−0.285***	(6.52)	−0.299***	(7.80)
season ending: 2018	−0.292***	(6.89)	−0.297***	(6.29)	−0.288***	(6.95)
season ending: 2019	−0.218***	(5.05)	−0.228***	(4.81)	−0.237***	(5.55)
Constant	13.417**	(156.92)	13.154**	(140.47)	13.376**	(123.51)
Observations	790		790		790	
adj-R^2	0.717		0.705		0.714	
Aic	32.769		114.374		42.362	
root MSE	0.239		0.251		0.240	

club dummies were also included in estimation; absolute t statistics in parentheses;.
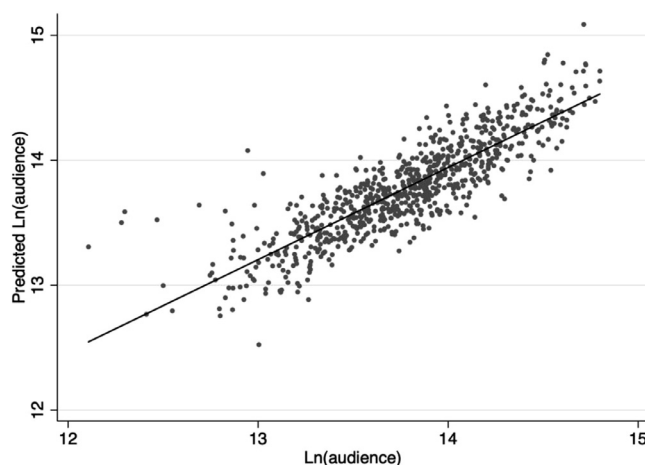
* $p < 0.1$,.
** $p < 0.05$,.
*** $p < 0.01$.

Fig. 3. Observed versus fitted values for Ln(Audience) (Model 1); the diagonal line is the $x = y$ line, shown for reference.

mates on the match significance variables, and in particular allows reasonable inference that there is some attraction to matches significant for European qualification. Amongst control variables, the coefficient estimate on 'weekday' changes from significantly negative in the programme audience equation to essentially zero in the match audience equation. This is despite the mean duration of both pre- and post-match content being more than twice as long

for weekend than for weekday matches: weekend programme audience data should be pulled down more through being diluted by non-match content. The result implies that there is much less propensity for audience members to view pre- and post-match content on a weekday evening. This is plausible given that, for many, the pre- and post-match periods come soon after work and before bedtime respectively. For the match itself, viewership does not seem to vary between the weekend and a weekday evening.

5.2. Player quality

Our measure of quality, which is average player rating across the two starting elevens, is strongly significant as a predictor of audience size. But, despite attracting a seemingly large estimated coefficient, the effect size is modest (relative to the influence of club brands, to be discussed below) though far from trivial. A match featuring a group of 22 starting players which had an average player rating one standard deviation above rather than one standard deviation below the mean would increase expected audience size by about 11%. In unreported experimentation, we tested for superstar effects by including an additional variable, the rating of the highest rated player. This variable proved decisively non-significant and its presence made minimal difference to other covariate estimates.

Buraimo and Simmons (2015) also report that player quality matters and infer that restrictions on player recruitment might be damaging, an even more relevant issue now with immigration restrictions following the United Kingdom's withdrawal from

the European Union. On the other hand, they may have been too hasty in their conclusion. Results where the player quality measure is significant demonstrate that British viewers are selective over which matches to view. The data cannot show how their behaviour might change if the average talent level in the League were lowered uniformly. They might continue to watch the same number of matches and continue to choose amongst them according to relative talent levels across matches.

We reviewed whether the use of our metric for player quality had made a material difference to findings. In three preceding studies of television demand for EPL football (Buraimo & Simmons, 2015; Forrest et al., 2005; Scelles, 2017), the alternative metric of 'combined relative wages' of the two clubs had been employed and this metric also featured in Caruso et al. (2019). Table 2, column 3, shows results from estimating our preferred model with combined relative wages substituted for average player rating. Had we been content with this variable, we would have concluded that viewers were unresponsive to player quality. Our assessment of the effect sizes from the match significance variables would also have been different. So introduction of our metric did indeed change findings in a substantive way.

Nevertheless, for all its imperfection, we were curious as to the complete failure of the combined relative wages variable to account for variation in audience size. It is plausible that the measure is at least positively correlated with whatever is meant by player ability (in our data set, the correlation coefficient between average player rating and combined relative wages was +0.655) and the preceding studies found a role for it in their modelling (for earlier periods than ours). In unreported regression, we re-estimated with combined relative wages as the player quality variable but with club dummies omitted. Now the combined relative wages measure was strongly significant. So it appears to be standing as a proxy for club dummies. Our interpretation is that, over the data period we analyse, there was such stability across seasons in the distribution of clubs' relative spending on wages that the information in the wages measure will have been collected in the coefficient estimates on the club dummies in our preferred equation. Recall that a weakness of the wages measure is that it is invariant whenever in the season a particular match occurs. The advantage of the player ratings measure is that it can exploit information readily available to fans concerning the actual ability of the players currently available to play and evolves over time. For example, it can represent a situation where a club has struck unusually well- or ill-judged contracts with new players such that actual rather than (wrongly) assumed player ability is measured. Likewise it can reflect information about changed personnel during a season, such as when a club has hired or sold important players or when a key player is lost to long-term injury. The additional variability allows viewer preferences for quality to be teased out and separated from the popularity of clubs. We therefore recommend that a metric of this type should be employed in future research.

5.3. Outcome uncertainty

Our outcome uncertainty measure was decisively non-significant. In case this finding concealed non-linear preferences, we experimented also with using a spline for outcome uncertainty but could identify no part of the range where the relationship had a non-zero slope. Our result is therefore *inconsistent* with the uncertainty of outcome hypothesis.

We are far from alone in failing to uncover evidence that, in competitions as currently constituted, viewers' decisions are influenced by how well-balanced a particular fixture is. Budzinski and Pawlowski (2017) noted that studies modelling the size of television audiences in sport "struggle in providing clear evidence for the importance of short-term uncertainty".

There is some previous work on television demand which does claim support for the uncertainty of outcome hypothesis. However, we are sceptical over whether that is what the relevant papers in fact established. Cox (2018) represents outcome uncertainty by including seven bands of 'home win probability' (as implied by bookmaker odds), using the band 5.9% to 17.6% as reference. The paper reports that the coefficient estimate on the band 35.9% to 45% was statistically significant (though none of the other bands had a significant impact). Now it is true that a home win probability around 40% would (once the draw probability was accounted for) indicate a finely balanced match. However, the results table indicates significance only at the 10% level. Further, in an alternative specification, the paper enters home win probability as a quadratic and neither component is significant even at 10% (whereas the uncertainty of outcome hypothesis would predict an inverted-U shape). While the paper claims to provide support for the uncertainty of outcome hypothesis, our reading of it is that the evidence offered points in the other direction.

Other authors have investigated whether consumers' tastes for uncertainty of outcome may have varied over time. Buraimo and Simmons (2015) and Schreyer et al. (2016) interacted uncertainty of outcome (from bookmaker odds) with season dummies. The first paper, which was on EPL viewing figures, reported that the outcome uncertainty variable was non-significant but the first two of the eight interaction terms were negative and significant (at 10%). It suggested that this pattern (significance in the first two years only), may have reflected an evolution of preferences away outcome uncertainty. However, this ignores multiplicity. There were eight tests on interaction terms and two were significant at 10%. With an appropriate adjustment in the p-value, to account for multiplicity, all the interaction terms, as well as the outcome uncertainty variable itself, would have been non-significant at 10%. Our interpretation of their evidence is therefore that there was no support in their data for the idea that viewers were responsive to uncertainty at any point in their data period.

Schreyer et al. (2016) also claimed to detect changing preferences. However, effect sizes were very small and neither the uncertainty variable itself nor the corresponding interaction terms were significant in modelling for Cup matches. In any case, we would caution against assuming that changing significance of interaction terms across successive seasons (or competitions) reflects changes in tastes. The competitive balance of a competition may vary from season to season and a significance threshold may or may not be achieved depending on the variability of the data.

Our finding that the relevance of outcome uncertainty to the prospective audience cannot be discerned from the data, is similar to that in much of the literature on both stadium and television demand. We believe that the result here is more credible than in many studies. Tainsky and Winfree (2010) pointed out that key variables in sports demand studies- player ability, short-term outcome uncertainty (here match outcome uncertainty) and medium-term outcome uncertainty (here match significance)- can be closely connected with each other and that it is therefore important to have a fully specified model if correct inference is to be possible. Some preceding work has failed to include measures of match significance or has represented it and other concepts inadequately.

Given that here and in prior literature, there is no clear support for the outcome uncertainty hypothesis, it might be questioned whether the concept deserves its central place in academic analysis of sports leagues.

5.4. Match significance

Two of the three match significance variables are highly statistically significant in our preferred model (and indeed in the alter-

native models in Table 2). Interest in the league championship is shown to be intense. A match with the highest championship significance in our data set would be expected to attract an aggregate audience size 96% higher than a match with no implications at all for end-of-season prizes but with the same clubs, players, etc. (the proportionate change in audience size for a log-linear model is calculated as $(e^\beta - 1)$). Matches with strong implications for Champions League qualification have weaker but still positive potential for attracting additional viewers although the variable is only borderline significant ($p = 0.051$).

At the other end of the table, a match with the highest relegation significance in our data set would be expected to raise audience size by 54% compared with a match of no importance for seasonal outcomes but otherwise similar characteristics. The relevance of relegation as a driver of demand appears to be understood by broadcasters. When the top division of English rugby union proposed to suspend relegation because of coronavirus disruption, BT Sport argued that it would then not gain full commercial value from its contract (Hunt, 2021). However, it should be noted that the absolute increase in viewership for matches in the EPL will be much lower for a relegation-significant than for a championship-significant match since it is likely to feature lower player quality and less popular clubs, such that the proportionate increase will be applied to a relatively low baseline figure.

Our measure of match significance was constructed by adding together the significance measure for each club. We recognised that this might be too strong a restriction and therefore we re-estimated all three models, including not only our aggregate measures of match significance but also the corresponding differences between the individual measures of significance for each club. For each model, none of the three additional terms was statistically significant.

The findings on match significance from our modelling are much more clear-cut than in preceding work on stadium and television demand for football, from which no evidence supporting even slight consumer interest in relegation or European qualification has emerged (Budzinski & Pawlowski, 2017). We believe that this literature has failed to pick-up the importance of seasonal outcomes to viewers because papers have not used appropriate metrics. That match significance matters is intuitively unsurprising because major team sports are nowadays almost always organised as league competitions. The alternative, followed in rugby union for much of its history (but now abandoned), would allow clubs to organise stand-alone matches with other clubs of similar standing. This should appeal to big market clubs to the extent that they could play only with each other, with high quality well-balanced teams on the field. That they instead choose to form a league suggests that they recognise that consumers will be more attracted by a structure with prizes at the end because the allocation of these becomes a source of interest in itself. Further, prizes incentivise intensity of effort, another reason for viewers to be drawn to 'significant' matches. Our results confirm the importance of the league structure and the creation of leagues within leagues (for example, for European qualification) for maximisation of audience. Outside Europe, while the American model of sport eschews relegation by holding to a closed-league format, the pressure to create match significance is still reflected in the idea of play-offs and the increased number of play-off places over time.

Past attempts to inform decisions on issues like how television revenue should be shared have typically been framed around 'competitive balance'. Even if match outcome uncertainty mattered, the link between it and competitive balance, which in the limit comprises equal talent at each club, would not be straightforward. As in other football competitions, home advantage in the EPL is an important factor in results: when we fitted the Maher model, the home advantage parameter was (for almost all seasons) about 1.2,

which is to say that home teams score goals at a rate 1.2 times higher than away teams. It follows that, if all teams had the same ability, match outcome uncertainty would still not be maximised since home advantage would mean that the home team would still have an appreciably higher win probability than the visitor (Forrest & Simmons, 2002, found that attendances tended to be particularly high at well-balanced matches where the superior quality of a visiting team was offset by home advantage; such matches would not be in the schedule if talent were equally distributed across clubs). However, our results show that match-level outcome uncertainty may not be important to viewers whereas interest in seasonal outcomes appears high. We would therefore argue that discussions on league format might give greater consideration to issues around match significance.

Budzinski and Pawlowski (2017) speculate that it might be less relevant to a league to 'improve' its overall competitive balance than to try to create intense competition for the championship between an oligopoly of well-endowed clubs at the top of the league and to have another cluster of clubs to serve as candidates in an intense struggle to avoid relegation. They argue that the distribution of financial and playing strength in the middle of the table is less important and therefore focusing on any index of overall competitive balance is misplaced. Our results support this contention as a possibility though it would also need to be taken into account that the distribution of talent within the mid-table cluster would affect the risk of a mid-table club being drawn into the competition for either European or relegation places, affecting the number of 'significant' matches.

The majority of income accruing to clubs in the EPL is generated by collective selling of broadcasting contracts and so the League has the policy instrument to bring about a revenue distribution which would move things in the direction advocated by Budzinski and Pawlowski (2017). But revenue redistribution implies talent redistribution, which would also impact audience size through the player quality variable. Simulation of different scenarios incorporating a television audience model such as we present here is recommended to inform future decisions by the League.

5.5. Controls

The season dummies present a story of declining demand, which might not augur well for future revenue growth. The pattern of viewership during a season is as one would expect, with audiences substantially higher in the winter months, November to February. A novel feature of our model is the inclusion of a dummy to capture the effect of the Christmas holidays. During this busy period of the English season, audience size appears to be about 10% higher than if the same matches had been scheduled for a regular weekday (or weekend). This may validate the practice of continuing through what would be a midwinter break in much of Europe. Quantification of the effect should allow a rational assessment, with the benefit balanced against perceived costs, such as higher risk of player injuries where there is no sustained period of rest during the August-to-May season.

Which broadcasting outlet shows a game proves important. In the first season, a match shown on BT Sports, a new entrant, was predicted to attract an audience only 46% of that which Sky Sports would be expected to achieve for the same fixture. By the last season, BT had achieved greater market penetration and was reaching an audience about 60% of Sky's. Splitting the contract between two broadcasters was imposed on the EPL by a ruling made under European competition law. Butler and Massey (2019) argued that having a second broadcaster actually hurt consumers by raising the cost of viewing. Here we see an apparent cost measurable in terms of lost viewing time.

Table 3

Club dummy coefficient estimates, with estimated effect size. AFC Bournemouth is the reference club.

	coef.	|t|	marginal effects
Manchester United	0.560***	10.89	0.750
Liverpool	0.549***	11.22	0.731
Arsenal	0.363***	7.51	0.438
Chelsea	0.323***	6.32	0.382
Tottenham Hotspur	0.297***	5.86	0.346
Manchester City	0.231***	4.56	0.260
Everton	0.231***	5.09	0.259
Newcastle United	0.224***	4.84	0.251
Queens Park Rangers	0.158**	2.27	0.171
Leicester City	0.152***	2.70	0.164
Southampton	0.147***	3.04	0.158
West Ham United	0.135***	2.83	0.144
Sunderland	0.118**	2.10	0.126
Crystal Palace	0.081*	1.78	0.085
Watford	0.078*	1.67	0.081
Norwich City	0.062	0.87	0.064
Huddersfield Town	0.060	0.74	0.062
Hull City	0.060	0.85	0.062
Middlesbrough	0.021	0.24	0.021
Swansea City	0.020	0.33	0.020
Stoke City	0.018	0.29	0.018
West Bromwich Albion	0.016	0.31	0.016
Burnley	0.010	0.20	0.010
Brighton and Hove Albion	0.002	0.02	0.002
Wolverhampton Wanderers	−0.056	0.43	−0.055
Fulham	−0.103	0.88	−0.098
Cardiff City	−0.111	1.35	−0.105

$*\ p < 0.1$,.
$**\ p < 0.05$,.
$***\ p < 0.01$.

Perhaps the most revealing results on controls are those on club dummies. Table 3 displays results from this part of the regression equation for our preferred model. Bournemouth was the reference club. They show the power of brands in football. Effect sizes sometimes dominate any contribution to audience size from our focus variables. As in Spain (Pérez et al., 2017), two clubs stand out in terms of appeal to a national audience. If Liverpool *or* Manchester United were substituted for Bournemouth, the 'brand effect' alone would be predicted to raise match audience size by about 75%. But, unlike Spain, there is a clear hierarchy of several other clubs with substantially higher drawing power than clubs similar to Bournemouth, all but one located in either Northern England or London. The strongest (Arsenal, Chelsea) have achieved major success relatively recently but some with still significant drawing power, e.g. Everton, Newcastle and Sunderland, continued to be popular despite decades of weak performance. Manchester City was the dominant club in our period but its success was a new phenomenon and evidently it had not (yet?) accumulated the appeal of Liverpool and Manchester United, which both performed under-par in most of the seasons we cover. The considerable variation in brand appeal across clubs points to one weakness of a relegation regime. Sporting performance may lead to the loss of clubs with significantly greater audience appeal than their replacements. Relegation of Newcastle (2016) and Sunderland (2017) will have dented aggregate audiences. More generally, by controlling carefully for player quality, outcome uncertainty and match significance, we have demonstrated how important brands are in English football and how enduring their ability to bring in audiences can be.

The large disparity between clubs in terms of television drawing power, illustrated here, helps explain continuing speculation over whether the stronger clubs may form a breakaway league or use the threat of one to force restructuring to a regime where they would take a higher share of revenue from collec-

tively sold television rights. In 2020, Liverpool and Manchester United proposed, but failed to secure, a League reorganisation which would give nine 'core' members, including the top seven clubs in terms of brand appeal identified in our analysis, protection from relegation and the right to sell television rights individually for a number of matches (Agini & Germano, 2020). The six clubs with the strongest television brands according to our results were the English clubs which attempted to form a (quickly aborted) pan-European Super League (Ahmed & Massoudi, 2021). Continuing tension must be expected given that, according to our results, some clubs contribute disproportionately to total audience size and therefore to the considerable revenue from broadcasting.

6. Concluding remarks

Our motivation in this paper was to address weaknesses we found in the literature modelling the size of television audiences for football matches, with the sport's most prominent league as the template. Addressing this problem is highly relevant and important both for the football industry, which needs to understand the market for televised matches if rights values are to be maintained, and for broadcasters, who commit substantial resources to procuring rights.

The weaknesses identified in the prior literature included an inappropriate definition of audience size, and inaccurate, misleading variables for representing player quality and match significance. Employing metrics used in sport analytics corrected these weaknesses and provided new, different insights into what draws audiences to view football matches. The results of general interest which stand out (and which future research might attempt to confirm, or not, for other major European football leagues) are the lack of support for the uncertainty of outcome hypothesis considered at the match level and, conversely, the importance to viewers of seasonal outcomes. Incidental results reveal the substantial differences between clubs in their ability to attract audiences, even controlling for player quality.

The implications of our findings are not limited to the academic understanding of what makes one football match more attractive than another. The very structure of a league and how it should be organised has long been influenced by an intense focus on competitive balance and outcome uncertainty. But our new findings on match significance, which, once properly measured, employing contemporary developments in sport analytics, proves to be important to the audience, points to new factors to be considered when looking to what competition formats would be most appealing to consumers. Further, companies considering sponsorship of teams or leagues, or advertising during matches, on television or at the stadium, need to consider the likely audience size to which their product will be exposed.

For sports leagues below the rank of the most elite, further research focused on attendance demand, their major revenue source, should use similar metrics as here to assess the preferences of the stadium audience and again use results to simulate possible reform in how competitions are organised.

Our results also point to a need to scrutinise published data for television audiences across a wider range of programmes than just for sport. We found material differences in audience size when only the event itself was included in the average-per-minute viewership. Other programme types, such as coverage of election results and royal events, also include 'build up' segments (and most programme types have advertising segments) where the audience size may be significantly lower, making comparisons across programmes and their commercial value problematic without using more detailed minute-by-minute data.

References

Agini, S., & Germano, S. (2020). Premier League proposals raise suspicions over US owners' motives. *Financial Times.* Accessed 28/03/2021 https://www.ft.com/content/22ffc8bb-81d8-4d8d-86f3-2b4c70b38ccb .

Ahmed, M., & Massoudi, A. (2021). Leaked Super League plans reveal goal of US-style football finances. *Financial Times..* Accessed 12/06 https://www.ft.com/content/e80299a4-8012-447a-8512-c24e149304b1 .

Bergmann, A., & Schreyer, D. (2019). Factors that shape the demand for international football games across different age groups. *International Journal of Sport Finance, 14*(1), 13–23. https://doi.org/10.32731/IJSF.141.022019.03.

Budzinski, O., & Pawlowski, T. (2017). The behavioral economics of competitive balance: Theories, findings, and implications. *International Journal of Sport Finance, 12*(2), 109–122.

Buraimo, B., & Simmons, R. (2015). Uncertainty of outcome or star quality? Television audience demand for English Premier League football. *International Journal of the Economics of Business, 22*(3), 449–469. https://doi.org/10.1080/13571516.2015.1010282.

Butler, R., & Massey, P. (2019). Has competition in the market for subscription sports broadcasting benefited consumers? The case of the English Premier League. *Journal of Sports Economics, 20*(4), 603–624. https://doi.org/10.1177/1527002518784121.

Caruso, R., Addesa, F., & Di Domizio, M. (2019). The determinants of the tv demand for soccer: Empirical evidence on Italian Serie A for the period 2008-2015. *Journal of Sports Economics, 20*(1), 25–49. https://doi.org/10.1177/1527002517717298.

Coates, D., & Parshakov, P. (2021). The Wisdom of Crowds and Transfer Market Values (April 2, 2021). Available at SSRN: https://ssrn.com/abstract=3818236 or http://dx.doi.org/10.2139/ssrn.3818236.

Coleman, B. J. (2012). Identifying the "players" in sports analytics research. *Interfaces, 42*(2), 109–118. https://doi.org/10.1287/inte.1110.0606.

Cox, A. (2018). Spectator demand, uncertainty of results, and public interest: Evidence from the English Premier League. *Journal of Sports Economics, 19*(1), 3–30. https://doi.org/10.1177/1527002515619655.

Dang, T., Booth, R., Brooks, R., & Schnytzer, A. (2015). Do TV viewers value uncertainty of outcome? Evidence from the Australian Football League. *Economic Record, 91*(295), 523–535. https://doi.org/10.1111/1475-4932.12206.

Deloitte (2020). *Home Truths: Annual Review of Football Finance, 2020*, June 2020. https://www2.deloitte.com/uk/en/pages/sports-business-group/articles/annual-review-of-football-finance.html Accessed 6 July 2021.

Dixon, M. J., & Coles, S. G. (1997). Modelling association football scores and inefficiencies in the football betting market. *Journal of the Royal Statistical Society, Series C (Applied Statistics), 46*(2), 265–280. https://doi.org/10.1111/1467-9876.00065.

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. https://doi.org/10.1111/1467-9884.00314.

Forrest, D., Simmons, R., & Buraimo, B. (2005). Outcome uncertainty and the couch potato audience. *Scottish Journal of Political Economy, 52*(4), 641–661. https://doi.org/10.1111/j.1467-9485.2005.00360.x.

Goddard, J., & Asimakopoulos, I. (2004). Forecasting football results and the efficiency of fixed-odds betting. *Journal of Forecasting, 23*(1), 51–66. https://doi.org/10.1002/for.877.

Hunt, H. (2021). RFU faces 'strong opposition' on relegation vote from BT Sport. *Insider Sport*. Accessed 28/03/2021 https://insidersport.com/2021/02/10/rfu-faces-strong-opposition-on-relegation-vote-from-bt-sport/ .

Jennett, N. (1984). Attendances, uncertainty of outcome and policy in Scottish League football. *Scottish Journal of Political Economy, 31*(2), 176–198. https://doi.org/10.1111/j.1467-9485.1984.tb00472.x.

Kharrat, T., McHale, I. G., & Peña, J. L. (2020). Plus–minus player ratings for soccer. *European Journal of Operational Research, 283*(2), 726–736. https://doi.org/10.1016/j.ejor.2019.11.026.

Liu, G., Luo, Y., Schulte, O., & Kharrat, T. (2020). Deep soccer analytics: Learning an action-value function for evaluating soccer players. *Journal of Data Mining and Knowledge Discovery, 34*, 1531–1559. https://doi.org/10.1007/s10618-020-00705-9.

Macdonald, B. (2012). Adjusted plus–minus for NHL players using ridge regression with goals, shots, Fenwick, and Corsi. *Journal of Quantitative Analysis in Sports, 8*(3), 1–24. https://doi.org/10.1515/1559-0410.1447.

Maher, M. J. (1982). Modelling association football scores. *Statistica Neerlandica, 36*(3), 109–118. https://doi.org/10.1111/j.1467-9574.1982.tb00782.x.

McHale, I. G., Scarf, P. A., & Folker, D. E. (2012). On the development of a soccer player performance rating system for the English Premier League. *Interfaces, 42*(4), 339–351. https://doi.org/10.1287/inte.1110.0589.

Pérez, L., Puente, V. V., & Rodríguez, P. (2017). Factors determining TV soccer viewing: Does uncertainty of outcome really matter? *International Journal of Sport Finance, 12*(2), 124–139.

Rosenbaum, D.T. (.2004). Measuring how NBA players help their teams win. https://www.82games.com/comm30.htm/Accessed 11 October 2020.

Scarf, P. A., Parma, R., & McHale, I. G. (2019). On outcome uncertainty and scoring rates in sport: The case of international rugby union. *European Journal of Operational Research, 273*, 721–730. https://doi.org/10.1016/j.ejor.2018.08.021.

Scarf, P. A., & Shi, X. (2008). The importance of a match in a tournament. *Computers & Operations Research, 35*(7), 2406–2418. https://doi.org/10.1016/j.cor.2006.11.005.

Scelles, N. (2017). Star quality and competitive balance? Television audience demand for English Premier League football reconsidered. *Applied Economics Letters, 24*(19), 199–203. https://doi.org/10.1080/13504851.2017.1282125.

Schilling, M. F. (1994). The importance of a game. *Mathematics Magazine, 67*(4), 282–288. https://doi.org/10.1080/0025570X.1994.11996232.

Schreyer, D., Schmidt, S. L., & Torgler, B. (2016). Game outcome uncertainty and television audience demand: New evidence from German football. *German Economic Review, 19*(2), 140–161. https://doi.org/10.1111/geer.12120.

Schreyer, D., & Torgler, B. (2018). On the role of race outcome uncertainty in the TV demand for Formula 1 Grands Prix. *Journal of Sports Economics, 19*(2), 211–229. https://doi.org/10.1177/1527002515626223.

Štrumbelj, E. (2014). On determining probability forecasts from odds. *International Journal of Forecasting, 30*(4), 934–943. https://doi.org/10.1016/j.ijforecast.2014.02.008.

Tainsky, S., & Winfree, J. A. (2010). Short-run demand and uncertainty of outcome in Major League Baseball. *Review of Industrial Organization, 37*(3), 197–214. https://doi.org/10.1007/s11151-010-9267-3.

Van Reeth, D. (2019). Forecasting Tour de France TV audiences: A multi-country analysis. *International Journal of Forecasting, 35*(2), 810–821. https://doi.org/10.1016/j.ijforecast.2018.06.003.

Van Reeth, D. (2020). Pitfalls in understanding reported TV Audiences for live sports broadcasts. Working paper Faculty of Economics and Business, KU Leuven, 30 pgs.

Wright, M. B. (2009). 50 years of OR in Sport. *Journal of the Operational Research Society, 60*, s161–s168. https://doi.org/10.1057/jors.2008.170.