Department of Economics

Economic Analysis Research Group (EARG)

University of Reading

Separating the crowds: Examining home and away attendances at football matches

by Brad Humphreys, J. James Reade, Dominic Schreyer and Carl Singleton

Discussion Paper No. 2022-11

Department of Economics
University of Reading
Whiteknights
Reading
RG6 6EL
United Kingdom

www.reading.ac.uk

Separating the crowds: Examining home and away attendances at football matches*

Brad Humphreys† West Virginia University

J. James Reade‡ University of Reading

Dominik Schreyer§ WHU

Carl Singleton¶ University of Reading

November 19, 2022

Abstract

The number of people consuming sporting events has long interested economists. Although imperfect, it is a measure of the demand for a 'peculiar' type of good or service — the sporting event. It also provides some measure of the social pressure on individuals performing. That pressure can be supportive, but it can also contribute to negative outcomes like choking on the part of performers. The extent to which a crowd is supportive or otherwise, however, is not always clear. In this paper we introduce a novel dataset detailing reported numbers of away fans at matches in England over recent years. We spend time characterising the dataset, and considering potential uses for it. We find evidence suggestive of different preferences for home and away fans; public holidays are a much stronger driver for away fan attendance, as is a team's league position. For away fans, whether or not the team remains in contention for end-of-season prizes matters much more than for home fans, and away fans are attracted by the novelty of a fixture more than home fans. We find some evidence that the expected number of away fans may have a small impact on match outcomes.

JEL Classification: Z2, R42, D91.
Keywords: Sport, home advantage, attendance, demand.

1 Introduction

The number of people consuming sporting events has long interested economists. Although imperfect, it is a measure of the demand for a 'peculiar' type of good or service — the sporting event Borland and

*We were delighted to be asked to contribute to this volume in honour of Professor Stefan Kesenne. The contributions of Stefan Kesenne to the economic understanding of sport are vast, as are his contributions to the sporting understanding of economics. One particular aspect of this has been the understanding of the impact of competitive balance, or the uncertainty of outcome hypothesis. In Szymanski and Késenne (2010) and Késenne (2000b), he considered the impact of revenue sharing on competitive balance, in Késenne (2000a) the impact of a salary cap, and in Késenne (2015) the optimal level of competitive balance. At the heart of all of these analyses lies the fundamental question of the demand for sport — what drives people to consume professional sport, and hence maintain its existence?

†John Chambers College of Business and Economics, Department of Economics. 1601 University Ave., PO Box 6025, Morgantown, WV 26506-6025, USA; Email: brhumphreys@mail.wvu.edu

‡Department of Economics, University of Reading, j.j.reade@reading.ac.uk

§Center for Sports and Management (CSM), WHU - Otto Beisheim School of Management, dominik.schreyer@whu.edu

¶Department of Economics, University of Reading, c.a.singleton@reading.ac.uk

MacDonald (e.g. 2003). It also provides some measure of the social pressure on individuals performing (Garicano et al., 2005). That pressure can be supportive, but it can also contribute to negative outcomes like performer choking (Harb-Wu and Krumer, 2019). The extent to which a crowd supports a team or does not, however, represents an elusive outcome. In many contexts, for example many team sports in North America, attendance may safely be categorised as entirely supporting the home team. This is not necessarily the case in other settings. English football, where historically fans often travel significant distances to observe their team, represents one such setting.

Attendance figures for sporting events, and for English football matches in particular, are widely available, facilitating many analyses of the demand for attendance (Schreyer and Ansari, 2022). But attendance data are not without their limitations. Attendance numbers are subject to measurement error, are censored above and below, are rarely provided alongside any pricing data, often conflate tickets sold with actual attendance, and studies usually assume that all in attendance are supporters of the home team. In recent years a number of studies have exploited an increasing array of more detailed data sources to circumvent the limitations of attendance data.

This paper represents one extension. We use detailed data on the number of home and away team fans to address the research question: does it matter for observable match outcomes if a non-trivial proportion of the attendance vocally supports the visiting, or 'away', team? We also explore the motivations of fans to attend sporting events by better characterising their team allegiance.

Since the 1970s, 'home' and 'away' fans have been segregated into different sections of stadiums in the top four divisions of English football.[1] In this paper we introduce and analyze a novel data set detailing reported numbers of away fans at football matches in England. We spend time describing the data, and discuss potential uses. One such example might be an evaluation of the public resource impact of significant numbers of fans travelling long distances at particular points in the week; another might be the environmental impact of that travel. We then perform some exploratory analysis, focusing on some basic explanatory variables for attendances disaggregated into home and away contingents. We also analyze a basic question of concern to football fans, and one considered in the literature previously: to what extent does fan presence influence match outcomes? We consider a slight twist, by considering whether the likelihood of an *away* win is influenced by the number of *away* fans in attendance.

We find evidence suggestive of different preferences for home and away fans; public holidays are a much stronger driver for away fan attendance, as is a team's league position. For away fans, whether or not the team remains in contention for end-of-season prizes matters much more than for home fans, and away fans are attracted by the novelty of a fixture more than home fans. We find some evidence that the expected number of away fans may have a small impact on match outcomes.

In Section 2 the context of English football is outlined and the relevant previous literature is reviewed, in Section 4 the modelling methodology adopted is set out, in Section 3 our dataset and sources are introduced, in Section 5 results from the econometric estimations are presented, and Section 6 concludes.

[1]These 'top four divisions' constitute the 'Football League', although since 1992 the top division has been part of a separate entity called the Premier League. In the years since relegation out of the top four divisions of the Football League has been introduced (1986) then expanded (2003), segregation has been more readily introduced into the fifth division, what is commonly referred to as 'non-league' football.

2 Context and Literature

We analyze demand for sport in England; that is, the recorded count of spectators attending football matches, traditionally recorded for taxation and revenue-sharing reasons.[2] Over time, such attendance numbers, or 'gates', developed into the standard metric for the size and importance of football clubs, and hence a source of pride for clubs and their followers. Being a proxy for a club's reputation, it's also important to other stakeholders (e.g., in terms of reach for corporate ponsors), including investors - cf. Gimet and Montchaud, 2016). Clubs consistently announce the attendance number at matches; see Reade (2021) who models attendance data in England back to the nineteenth century.

Football's nineteenth century league origins in the densely populated North West and West Midlands regions of England, with their extensive rail networks, meant that from very early on followers of a club would travel to observe their team play at another stadium.[3] In the mid-1970s, however, following sharp increases in incidents of violence between fan groups, teams introduced segregation between home and away supporters in stadiums (Lowles and Nicholls, 2005). The increasing problem of fan violence, or hooliganism, in England led to more significant security measures, most notably the requirement that all top clubs had to have all-seater stadiums, rather than the historically common standing terraced areas. Away support at football matches, though segregated, remained strong throughout this period. In our database we have observations of multiple thousands of away supporters at matches played in the late 1980s and early 1990s.

In recent years, when announcing attendances both at the stadium during a match, and on social media, clubs commonly announce the number of away supporters in attendance. Attendances, and their size, have long been a central part of fan culture, with bragging rights at stake, and low attendances commonly a source of mockery from rival fans. It is likely that this drives the increase in the common knowledge of away supporter numbers, and in constructing the data set described in this paper, we exploit this increased detail in attendance numbers.

The economics literature primarily focussed on attendance numbers, not least as they represent revealed preferences by individuals regarding the purchase of a good or service. More often than not, though, price data do not exist. Reade (2020) analyzes turnstile data at one English football stadium that includes different prices paid by spectators, and Arnold (1991) make more general use of club receipts from matches. As Arnold (1991) note, these data are generally unavailable for most matches.

Another generally unavailable aspect of such aggregate attendance numbers is the proportion of those in attendance who have purchased a season ticket — a bundle of tickets for all home league matches played by club, generally purchased before the season begins. As such, there are many aspects of the composition of a crowd that may matter for outcomes, but cannot usually be measured. Another characteristic of crowd composition explored in Singleton et al. (2021) is the ease with which fans can attend; they considered Egyptian football, where for a period attendance was heavily restricted on security grounds to home fans vetted by authorities.

Attendance numbers have been used historically to consider the uncertainty of outcome hypothesis (Peel and Thomas, 1988, 1992; Coates and Humphreys, 2010, 2012; Cox, 2018; Schreyer and Ansari, 2022). Despite the clear value of attendance numbers for such investigations, attendances are nonethe-

[2]Technically we focus on England and Wales, as a small number of Welsh clubs have traditionally played in English leagues: Cardiff City, Swansea City, Newport County and Wrexham.

[3]For a good source of information on the development of away fan culture in the UK, see 'Travelling to support a football club is a habit and it's a habit that's been broken', *The Athletic*, January 23 2021 https://theathletic.com/2338893/2021/01/23/travelling-to-support-football-is-a-habit-and-its-a-habit-thats-been-broken/.

less limited in their utility for this purpose, since they are often censored above, and are subject to measurement error. More recently, television viewership has been considered (Buraimo and Simmons, 2009; Forrest et al., 2005; Buraimo and Simmons, 2015), alongside social media activity (Garcia-del Barrio and Reade, 2021).

Attendance numbers have also been used to understand more about the impact of social pressure on match outcomes, in particular in relation to the well-known home advantage in sport (Sutter and Kocher, 2004; Pettersson-Lidbom and Priks, 2010; Dawson and Dobson, 2010; Buraimo et al., 2010; Page and Page, 2010; Dohmen and Sauermann, 2016), with a particular focus of this coming with the absence of crowds during the Covid-19 Pandemic (Bryson et al., 2021; Endrich and Gesche, 2020; Cueva, 2020; McCarrick et al., 2021; Reade et al., 2022). The Covid-19 Pandemic also raised awareness of the role mass attendance events like football matches can contribute to the spread of an airborne virus (Stoecker et al., 2016; Cardazzi et al., 2020), with Olczak et al. (2020), Fischer (2022) and Alfano (2022) explicitly considering the spread of Covid-19 in the areas surrounding the stadiums of teams playing. Olczak et al. (2020) in particular consider the impact in the areas surrounding the stadiums of away teams, finding a similar effect to that in the areas surrounding the home team's stadium.

The public resource implications of large sporting events are not constrained only to the spread of viruses. Medical services are required, as are policing and stewarding services, diverting significant resources away from other uses when a game is taking place. The impact of many thousands of fans travelling around the country on weekends and weekday evenings also puts additional strain on transport networks.

It is also an environmental concern, increasingly, as football amongst other sectors recognises its need to become more sustainable. Wicker (2019) is an example of an increasing recognition of this, and the travel, often in individual cars, of football fans week by week during the season (10 out of 12 months of the year) must make a contribution.

3 Data

Data on aggregate attendance numbers are readily available from a number of sources. We collect them primarily from the websites footballwebpages.co.uk, soccerbase.com and 11v11.com. We merge additional information on the geolocation of each team's home stadium from worldfootball.net. We collect, and merge in, betting price information from football-data.co.uk and oddsportal.com.[4]

Our data on away attendances is collected from a range of sources. The main source is via social media. In recent years it has become very common that when an attendance is announced at a football match, the number of away fans are also announced. We have thus taken data from a range of Twitter accounts, including official accounts of clubs,[5] but also accounts that commonly report announced away attendance figures.[6] We have also taken data from club sources of attendances: Barnet FC, Torquay United FC, and Wycombe Wanderers FC have fan-operated websites containing extensive records of away attendances going back many years.[7]

[4]The former considers only league matches but generally goes back further, historically, than the latter, but the latter includes all cup competitions, and even friendly matches.

[5]Mansfield Town, Oldham Athletic, Stockport County, Grimsby Town, Wrexham, Luton Town and Aldershot. We are thankful to Mark Airey, an Oldham Athletic supporter who compiled many away attendances involving Oldham matches that feature in our dataset.

[6]d3d4football, d5football, and fanbanter.

[7]These websites are www.downhillsecondhalf.co.uk/ for Barnet, www.torquayfanstats.com/ for Torquay United and www.chairboys.co.uk/ for Wycombe Wanderers.

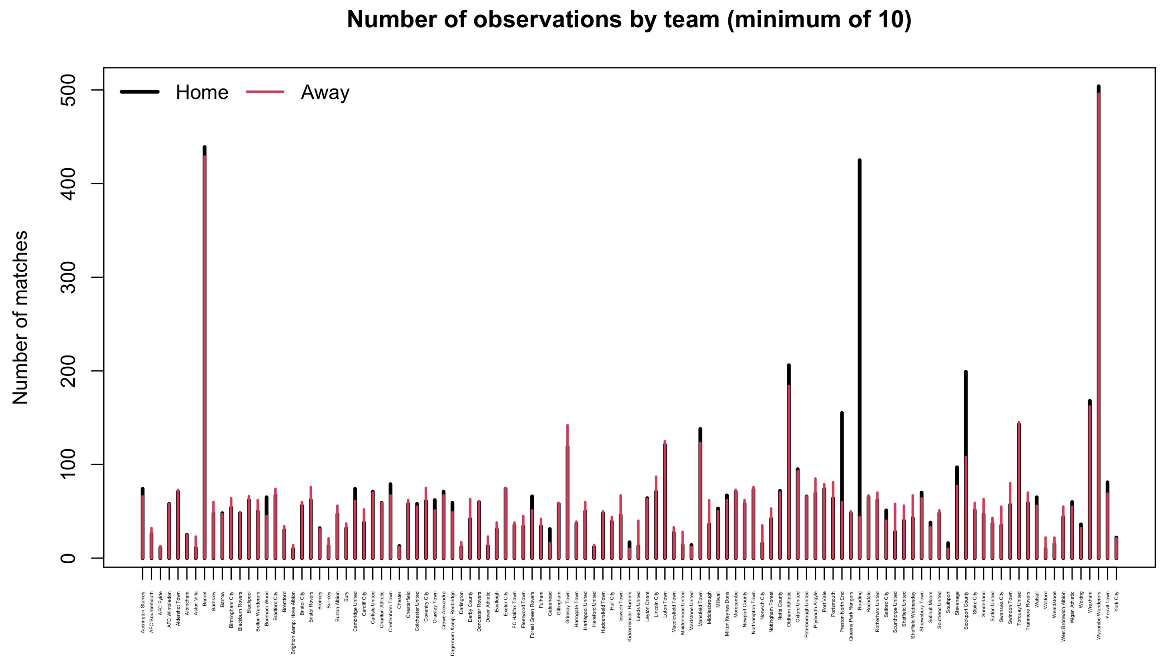

Figure 1: Distribution of the observations in the data across football clubs. The black lines are the number of matches in which that team is the home team, the red lines the number of matches in which that team is the away team.

Our second main source of data comes from a small number of football clubs themselves, who have generously provided detailed stadium records from home matches over many years, which include the number of away fans in attendance. These are Reading FC, Oldham Athletic AFC and Preston North End FC.[8] The spread of observations across different teams is displayed graphically in Figure 1; in this figure, the black bars are matches in which that team is playing at home, and red bars are matches in which that team is playing away. The majority of teams have around 50 observations at home and away, and the spread is quite broad, with the teams mentioned above constituting the spikes, with Barnet, Reading and Wycombe Wanderers the three largest spikes.

An important question when considering attendance data is the potential for measurement error. While numbers are commonly released by clubs, they are not necessarily the total number of people inside the stadium during a match. Historically, it has been believed that clubs understate attendance numbers, in order to minimise tax bills. Equally, football clubs will often wish to overstate attendance numbers, since attendance is a measure of the popularity and viability of a football club. Quite commonly, the number announced is the total number of tickets sold, while the actual number in the stadium, whilst collected, is not reported. Schreyer et al. (2019) has documented this 'no show' behaviour on the part of fans who buy tickets, but do not subsequently attend. It is not clear what to make of the potential for measurement error, given it has the potential to be both systematically upward, and downward biased. In this paper, we take the data at face value, hence assuming that, on average, the numbers are correct.

[8] The gate books for Preston North End were accessed at the National Football Museum's Archive at Deepdale in Preston.

Summary statistics are presented in Table 1. Our earliest observations of away attendances come from 1989 at Preston North End, and hence from 1989/90 onwards we collect information on all league matches. From 2005 onwards, we add non-league matches in the first level beneath the Football League (now called the National League).[9] As such, our dataset is, in total, over 120,000 matches with full attendance information.

Our information on away attendances is for around 7,500 matches, and hence covers about 6% of these matches. The bulk of these matches are in lower divisions; 1,313 are in the Championship (second division), 1,623 in League One (third division), 2,450 in League Two (fourth division), and 1,269 in the National League (fifth division). As such, the average total attendance is lower in these matches at (7,058+832=) 7,890 fans, relative to our overall mean of 8,732. Note that the minimum attendance, for home and away fans, is zero — these are matches during the Pandemic behind closed doors. Because we consider the log of attendance in our empirical models, these closed-door matches are excluded. The smallest away attendance in our estimation sample is seven people, which occurs twice.[10]

To calculate the distance in miles between the teams taking part in a match we use the geolocation of each team's stadium as provided either on the worldfootball.net website or a club's Wikipedia profile, and for population we use UK census data from 2011 on the local area in which a team's home stadium is located. On average, there is about 96 miles between teams taking part in matches in our dataset, and teams have a local population of about 232,000, albeit with a sizeable standard deviation (168,500).

Home teams win about 45% of the time in our dataset, and away teams 31%. The bookmakers, by and large, anticipated this, expecting the home team to win about 43% of the time, and the away team around 30%.

Table 1: Summary Statistics

Statistic	N	Mean	St. Dev.	Min	Pctl(25)	Pctl(75)	Max
Attendance	120,856	8,732.069	11,921.720	0	909	11,658	89,874
Home Attendance	7,526	7,064.661	7,074.519	0	2,367	8,988.500	70,870
Away Attendance	7,533	829.269	940.727	0	200	1,148	9,500
Home team pre-match implied probability (mean)	88,456	0.431	0.129	0.025	0.348	0.515	0.935
Away team pre-match implied probability (mean)	88,456	0.307	0.118	0.019	0.223	0.373	0.929
Home win (1/0)	157,538	0.446	0.497	0	0	1	1
Away win (1/0)	157,538	0.309	0.462	0	0	1	1
Distance (miles)	146,289	96.412	61.350	0	46.014	138.766	334.141

In Figure 2 we plot the attendances for matches where we have information on the away attendance, with the away attendance in red. The match attendances are plotted with the date on the horizontal axis, giving a sense of the chronological spread of our data. Our data from Preston North End is in the early 1990s, with a single observation in the late 1990s, and from the early 2000s matches for Reading, Barnet and Wycombe Wanderers are quite regular. The regular gaps are between seasons, with a large gap in 2020 into 2021 caused by the Covid-19 Pandemic, which forced almost all matches to take place without fans. The information from websites like Fanbanter, which covers the Championship down to the National League, appears in late 2018 at first, and hence the larger overall crowds are recorded from then.

In Figure 3 we give a sense of the proportionality of away crowds to home crowds by looking at

[9]The footballwebpages website lists attendances at a number of non-league levels, but below the National League these do not, in general, report away attendance numbers since fans are not segregated formally.

[10]These are both matches involving Stockport County, a professional team that dropped into regional football over our sample period before recovering in recent seasons.

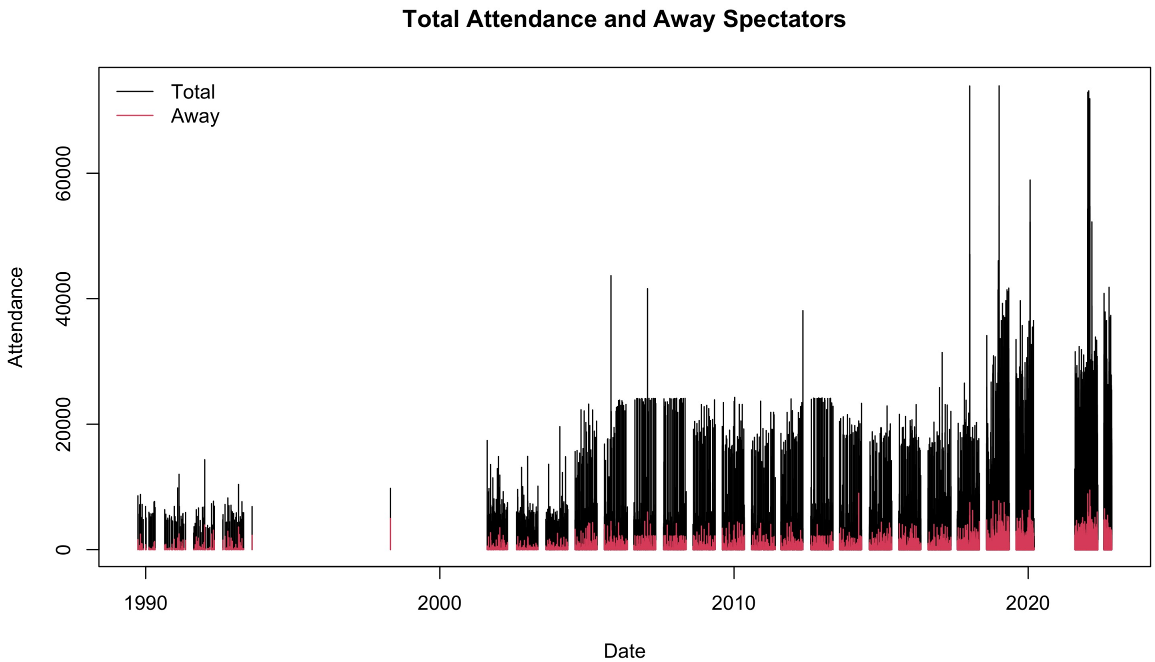

Figure 2: Attendances at English football matches where away attendance (in red) has been recorded. The matches are plotted against date on the horizontal axis.

the ratio of recorded away attendances to total attendances. The distribution of the ratio of the away attendance to total attendance is somewhat skewed, with a median of 9% and a mean of 12%, reflecting that in a small number of matches, the away crowd can dominate. Indeed, in 62 matches in our sample, there are more away fans counted in the stadium than home fans. In Figure 4 we plot the relationship between the distance an away team has travelled and the number of its fans at the game. The further a team travels, the fewer of its fans travel to attend. The unconditional marginal effect is about a 5% fall in away attendance for each 10 miles a team travels.

The matches with away attendance in the data represent contests in a number of different competitions. The majority are league matches (6,878) at various levels in the FA. The remainder of matches represent matches played in cup competitions like the FA Cup, the League Cup, the FA Trophy, and the Football League Trophy.

4 Methodology

The main purpose of this paper is to describe our novel data set, and to do so, we first estimate a predictive model for home and away attendances, in order to determine whether different factors motivate attendance decisions of different types of fans. We then use these predictive models in a second stage of regressions analyzing match outcomes. The rationale for this approach is to consider the extent to which the size of both home and away crowds, and the amount to which such sized crowds were expected, may affect match outcomes.

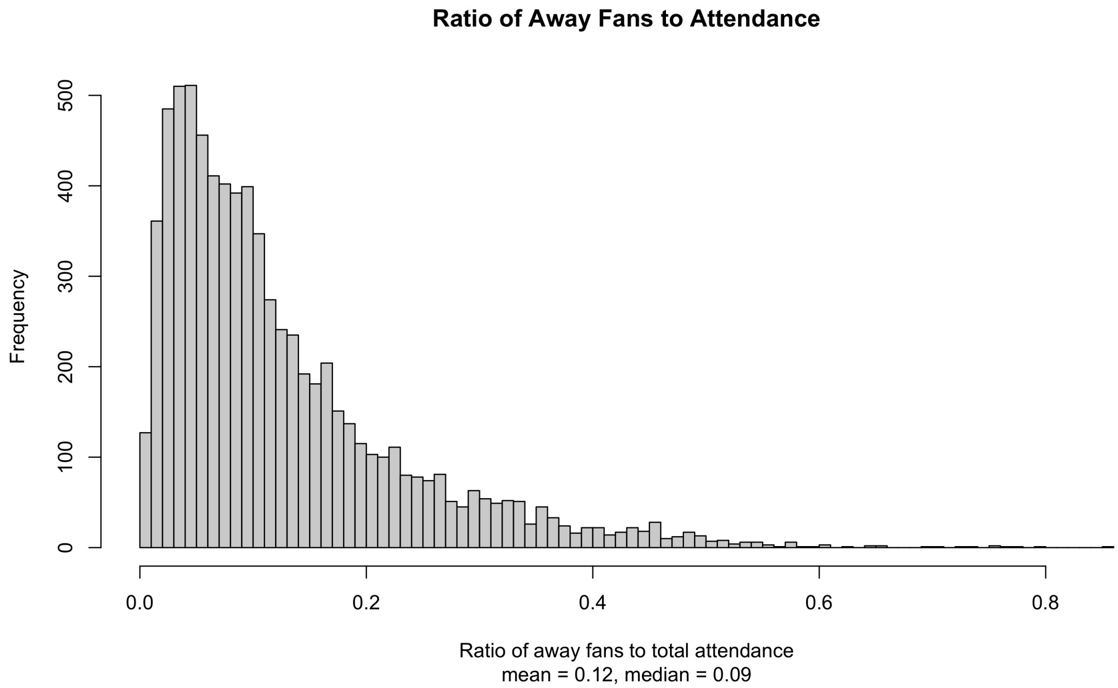

Figure 3: Distribution of the ratio of away attendance to total attendance for matches where the away attendance is known.

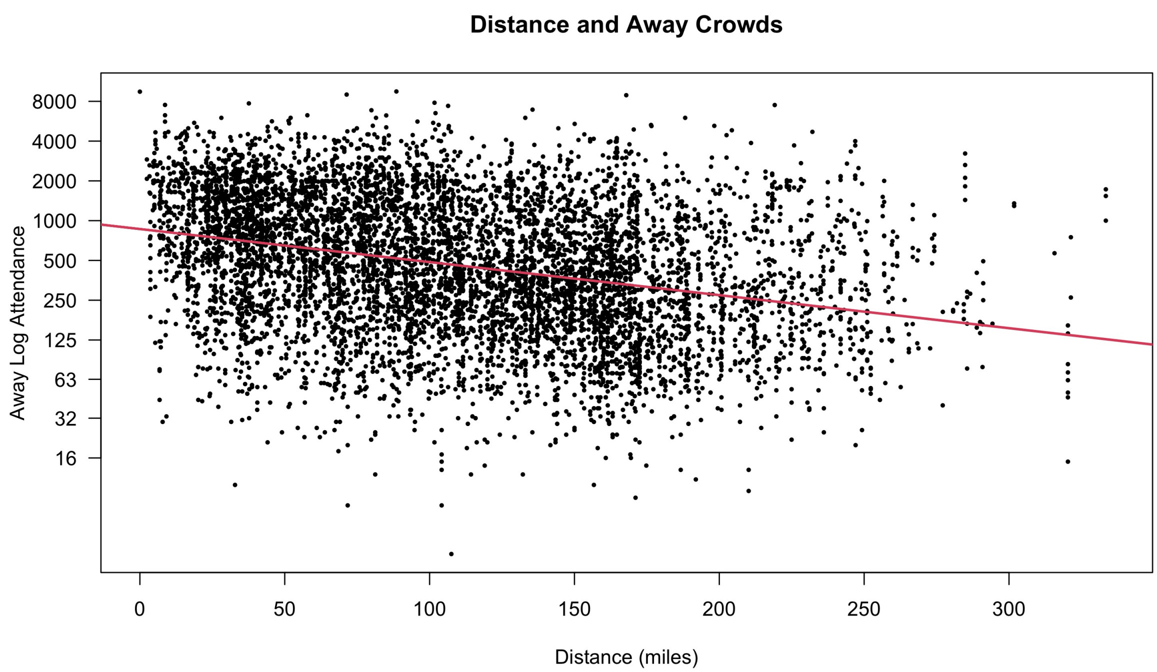

Figure 4: Scatter plot of distance a team travels and the number of visiting team fans attending.

We regress attendance on a number of commonly used, and intuitive, explanatory variables. We first consider variables that pertain to the circumstance of the match event: we include the distance between the two teams competing in the match, and include home and away team year fixed effects, as well as controlling for the day of the week, month of the year, and whether or not a match took place during a public holiday period. We have not included information on population estimates for the local area, as these do not change over the time period. Instead, to allow attendances to vary club-by club and year-on-year due to unobservable factors, we interact home and away team fixed effects with year fixed effects. Although we do not have information on prices or incomes, we factor in budgetary constraints by considering the length of time since a team last played a home or away match. Conventionally, fans buy season tickets, which allow entry to all home matches, and hence a number of home matches in succession may not matter so much for attendance decisions, but a number of away matches in close succession may have an impact on budgets.

Next, we consider variables that dictate match attractiveness. We firstly consider the likely outcome of the match, and how close it is anticipated to be. To do so, we use the Elo rating system (Elo, 1978). Elo ratings are intuitive in their design; each team has an Elo rating, which we denote R_i and R_j for team i and team j. The Elo prediction for such a match at time t, from team i's perspective, is:

$$E_{it} = \frac{1}{1 + 10^{(R_j - R_i)/400}}. \tag{1}$$

The prediction from team B's perspective (E_{jt}) is defined similarly, and are such that $E_{it} + E_{jt} = 1$. Note that (1) is a logistic curve with base 10, and hence is bound on the unit interval, so $0 \leq E_{it} \leq 1$. The Elo prediction, E_{it}, can thus be thought of as a measure of expectation regarding a match outcome; with a larger E_{it}, team i is expected to win. Then each team's Elo rating is adjusted once the outcome S_{it} is known, where:

$$S_{it} = \begin{cases} 1 & \text{if team } i \text{ wins,} \\ 0.5 & \text{if game is a draw/tie,} \\ 0 & \text{if team } i \text{ loses.} \end{cases} \tag{2}$$

Again, an equivalent variable S_{jt} can be defined, and $S_{it} + S_{jt} = 1$.

Then the adjustment is:

$$R_{it+1} = R_{it} + K(S_{it} - E_{it}). \tag{3}$$

Elo ratings have been commonly used in a range of sport economics papers (Carbone et al., 2016; Vaughan Williams et al., 2019). In addition to using E_{it} as a measure of the strength of the home team relative to the away team, we also define the following measure of *balance* in a match:

$$balance_{ijt} = E_{it}(1 - E_{it}) = E_{it}E_{jt}. \tag{4}$$

This measure achieves its highest value when $E_{it} = 0.5$, which is when both teams are exactly evenly matched.

We then further include three arguably more salient measures of a match's quality, all of which are league-table based. We include the league position of each team, the goals scored by teach team, and the number of points that that team is from 6th place. As our data are all from English divisions beneath the Premier League, 6th place usually dictates the lowest position in the table that allows a team to qualify for the end-of-season play-off tournaments that can result in promotion.[11] That this variable is calculated for teams both below 6th in the table (the measure is positive) and those above 6th (the measure is then negative).

These league-based measures exclude cup matches from consideration.[12] In principle, a team's league-based attributes can be copied across for cup matches. However, it is likely a different decision-making process generating attendance at cup matches (see e.g. Szymanski, 2001). In addition, season tickets for attendees do not cover cup matches and hence the economic decision and crowd composition is very different relative to league matches.[13] In this paper we focus on league matches.

We have neither income nor price information, but we can factor in some decision making on the part of fans in relation to budget constraints and other aspects of the attractiveness of the football match in question. We can calculate for the home and away team in a match how many days it has been since the most recent home or away match. As it might be anticipated that season ticket holders form much of the proportion of home spectators, then a concentration of home matches wouldn't be as problematic from a budgeting perspective as a concentration of away matches.

Our first regression model is thus, for a match between home team i and away team j at time t:

$$\begin{aligned} att_{ijt} = \beta_{0ijt} &+ \beta_1 distance_{ij} + \beta_4 E_{it} + \beta_5 balance_{ijt} + \beta_6 pos_{it} + \beta_7 pos_{jt} \\ &+ \beta_8 gscore_{it} + \beta_9 gscore_{jt} + \beta_{10} pts6th_{it} + \beta_{11} pts6th_{jt} \\ &+ \beta_{12} days.since_{it} + \beta_{13} days.since_{jt} + \beta_{14} years.since_{ijt} + e_{ijt}, \end{aligned} \tag{5}$$

where β_{0ijt} refers to our set of fixed effects, which are for each team each year, the day of the week, week of the year, and division that the match took place in. Team-year fixed effects are included because of the extensive changes that take place between seasons (season ticket purchases, playing squad recruitment) that can mean attendances, home and away, can be quite different.

We run this regression using the least squares method of estimation, with dummy variables for fixed effects, and we look at total attendances as well as home and away attendances separately. Our purpose in running this initial stage regression is to develop a predictive model that can then be utilised in our second stage.

We use the fitted (or predicted) values $\widehat{att}_{ijt}$ from (5), as well as the residuals, $\widehat{e}_{ijt} = att_{ijt} - \widehat{att}_{ijt}$, as an explanatory variable in a regression of a limited range of outcome variables. We do this to answer the question regarding whether "surprise" attendances matter, rather than attendances per se. The in-sample predicted values, $\widehat{att}_{ijt}$, we treat as the *expected* part of each attendance, and the residuals, $\widehat{e}_{ijt}$ we treat as the *unexpected*, or surprise, part of each attendance. We thus decompose each attendance.

We are unaware of any such similar approach to understanding the dynamics of outcomes in relation to crowds. It seems plausible that agents involved in a football match form an expectation regarding the likely attendance, and this may inform their preparations, pratically and mentally. It is possible that an unexpectedly large, or small, crowd, could disturb these expectations at a late stage in the planning

[11] In England's League Two, the top three teams are automatically promoted, with the four beneath qualifying for a play-off tournament, and in the two divisions beneath only the top team is promoted and the six beneath qualify for a play-off tournament, and hence seventh rather than sixth is the lowest position. In general, though sixth and seventh places are not separated by more than a few points.

[12] And note further that in English football, as with all European leagues and many worldwide, league and cup competitions occurs at the same time over the calendar, rather than a distinct, sequential league and cup/play-off competitions.

[13] In England and Wales, it is uniformly the case that season tickets do not include cup matches or away matches, though this may be different elsewhere in Europe. Season tickets will usually though afford the holder some priority in purchasing tickets for cup matches and away matches.

process for a football match.

We consider whether the simple home and away attendances, or their expected levels, or their unexpected components matter for whether or not away wins occur. It is conventional to consider the home advantage, but as our dataset allows insight into the number of away supporters in attendance at a match, it seems interesting to consider instead the likelihood of an away win. Our dependent variable is hence a binary variable taking the value one if an away win occurs.

Our second-stage regression models are thus:

$$\begin{aligned} awaywin_{ijt} = \gamma_{0ijt} &+ \gamma_1 att_{ijt} + \gamma_2 distance_{ij} + \gamma_3 pop_i + \gamma_4 pop_j \\ &+ \gamma_5 E_{it} + \gamma_6 balance_{ijt} + \gamma_7 pos_{it} + \gamma_8 pos_{jt} \\ &+ \gamma_9 gscore_{it} + \gamma_{10} gscore_{jt} + \gamma_{11} pts6th_{it} + \gamma_{12} pts6th_{jt} + v_{ijt}. \end{aligned} \tag{6}$$

$$\begin{aligned} awaywin_{ijt} = \gamma_{0ijt} &+ \gamma_1 \widehat{att}_{ijt} + \gamma_{1b} \widehat{e}_{ijt} + \gamma_2 distance_{ij} + \gamma_3 pop_i + \gamma_4 pop_j \\ &+ \gamma_5 E_{it} + \gamma_6 balance_{ijt} + \gamma_7 pos_{it} + \gamma_8 pos_{jt} \\ &+ \gamma_9 gscore_{it} + \gamma_{10} gscore_{jt} + \gamma_{11} pts6th_{it} + \gamma_{12} pts6th_{jt} + v_{ijt}. \end{aligned} \tag{7}$$

The first model, equation (6) uses *actual* home and away attendances, while the second, equation (7) decomposes actual attendance into our expected, and unexpected, attendance.

It is worth noting that the expected attendance we use is in an econometric sense expected; it is the prediction from a reasonably complex econometric model. It need not necessarily correspond to what any individual player, or any official, may consider to be the expected crowd. It may be that such agents use simple rule of thumb forecasting methods, but we are unable to model their expectations, naturally. We consider the econometric expected, and unexpect, attendance components to nonetheless be informative about some collective expectation of attendances.

5 Results

We present the results from estimating the attendance equation in (5) in Table 2, first, before presenting the away attendance equation (6) in Table 3. In the first column we estimate over all league matches for total attendance. In the second column we estimate over total attendance, but restrict attention only to matches for which we have data on away attendances. In the third and fourth columns we then estimate over the home and away attendances, respectively.

Distance exerts a negative influence on attendances, with the conditional marginal effect being smaller than the unconditional one reported earlier: about a 0.1% fall (compared to 0.5%) in attendance for every extra 10 miles between teams. This effect is driven mainly by the away following, as the final two column suggest. The effect on the home crowd is actually positive, but barely statistically significant and small, economically. The effect on the away crowd is slightly larger than the overall effect, and in the right direction: each 10 miles reduces the away crowd by 0.8%.

The relative strength of the two teams is insignificant in explaining vaiation in attendance, and while balance appears to matter for total attendances, it is insignificant for home and away attendances separately. This is primarily due to the greater noise in estimates based on the smaller sample of matches with away attendances.

The impact of league position, the most salient statistic for fans of a club, is again better considered separately. While in the first two columns the effect for both clubs is significant and in the direction to be expected (an extra league position adds 0.07% and 0.02% to the game for the home and away team respectively), the effect is larger for both sets of fans when considering them separately (0.8% and 1.3% per position, respectively). More goals scored by the home team attracts more fans to attend — both home and away fans.

The proximity of a club's current position from 6th place on the league table is also significant in explaining variation in attendances. Each extra point away from 6th decreases attendance by 0.06% or more, with home attendances being more affected by the distance of the home team (0.08%), and away attendances more affected by the away distance (0.2%).

By separating attendance into home and away segments, we see that fans are influenced by the characteristics of their own team. While characteristics of each team influences the total attendance, many of the coefficients on characteristics of the home and away team are larger in the third and fourth columns where we only consider the home and away fans in isolation. The coefficients on the Elo prediction are larger for home and away fans separately, and the league position of the away team has a larger coefficient for away fans than total fans, and the number of points from sixth place. There is undoubtedly a cross-team influence; the attractiveness of the opposition must be a factor in decisions by fans to attend matches of their team, and indeed Schreyer et al. (2019) find clear evidence of this in 'no show' behaviour of fans. But it is also intuitive that fans are most strongly influenced by the temporal characteristics of their own team.

Finally, timing factors affect attendance; if teams are faced with more matches in a short space of time, then attendances will be lower; each extra day since a team's most recent home or away match increases the attendance by around 0.1%. Additionally, away fans are more likely to travel to a stadium if it has been more years since their team last played at that venue.

Having estimated our first stage regression model, explaining attendances for both home and away fans, we can now use that information to consider how it might impact actual on-field outcomes.

Table 3 shows the model estimates along columns. The dependent variable in each case is a binary variable for an away win in a match. We report different versions of equations (6) and (7) across the six columns. Estimated standard errors are clustered at the home and away team level.

We firstly report without fixed effects or control variables, with column (1) estimating equation (6) and column (2) estimating equation (7). Here, we find that away wins are less likely with a larger home crowd, and more likely with a larger away crowd, which is consistent with the notion of home advantage that outcomes for a team are a function of the size of a supportive crowd.

In the second column we decompose home and away attendances into their *expected* and *unexpected* components, according to our first stage econometric model. The coefficient sizes on the two sets of fitted values, or expected attendances, are very similar to the coefficients on actual attendances in column (1). The coefficients on the residuals, or the unexpected attendance, are relatively small, and statistically insignificant. As such, it is the expected crowds that matter, rather than the unexpected parts.

In the third column we add fixed effects for home and away team interacted with a variables for individual seasons. When doing this, the impact of the actual home crowd becomes statistically insignificant, and of the opposite sign though similar sign. The impact of the actual away crowd shrinks to essentially zero, and becomes insignificant. As such, the strengths of teams at home and away across individual seasons explains the variation in away wins more than the size of the actual crowds attending matches. In the fourth column we further decompose the actual attendances into expected and unexpected components. Again, only the expected crowd appears to exert any significant impact on attendance, but the sign is different from before. Controlling for seasonal variation in the strengths of home and away

Table 2: Results from estimating the attendance equation (5).

	Dependent variable:			
	log attendance		log home attendance	log away attendance
	(1)	(2)	(3)	(4)
Distance	−0.001***	−0.001***	−0.0004***	−0.008***
	(0.00002)	(0.0001)	(0.0001)	(0.0002)
Holidays	0.071***	0.063***	0.061***	0.193***
	(0.004)	(0.015)	(0.013)	(0.041)
Elo Prediction	0.072	0.431	1.284***	−0.917
	(0.072)	(0.328)	(0.304)	(0.609)
Balance	0.003	0.382	−0.350	−1.095
	(0.074)	(0.411)	(0.379)	(0.858)
Position (H)	−0.006***	−0.004***	−0.003***	−0.003
	(0.0003)	(0.001)	(0.001)	(0.002)
Position (A)	−0.002***	−0.002***	−0.001**	−0.010***
	(0.0002)	(0.001)	(0.001)	(0.002)
Goals scored (H)	0.002***	0.001	0.001	0.001
	(0.0003)	(0.001)	(0.001)	(0.002)
Goals scored (A)	0.002***	−0.0001	−0.0003	0.001
	(0.0003)	(0.001)	(0.001)	(0.002)
Points from 6th (H)	−0.003***	−0.001	−0.001	−0.001
	(0.0003)	(0.001)	(0.001)	(0.003)
Points from 6th (A)	−0.001***	−0.003**	−0.003**	−0.007**
	(0.0003)	(0.001)	(0.001)	(0.003)
Days since last match (H)	0.0005***	0.001**	0.001***	
	(0.0001)	(0.0003)	(0.0002)	
Days since last match (A)	0.0001	0.0003		0.001***
	(0.0001)	(0.0003)		(0.0005)
Years since last match-up	0.0004***	0.001**	−0.00001	0.003***
	(0.0001)	(0.0003)	(0.0003)	(0.001)
Observations	101,519	6,873	6,873	6,873
R^2	0.987	0.982	0.986	0.939
Adjusted R^2	0.985	0.969	0.976	0.895
Residual Std. Error	0.187 (df = 91497)	0.153 (df = 3973)	0.142 (df = 3974)	0.370 (df = 3974)

Notes: Standard errors are clustered at the home and away team-year level. Fixed effects are included for home and away team years, the day of the week, week of the year, and division the match took place in.
*p<0.1; **p<0.05; ***p<0.01

teams, a larger home crowd reduces the likelihood of an away win.

In the fifth column we add in a set of control variables related to the measurable strengths of teams. These three controls include current league positions, Elo strengths, and goals scored and points from 6th in the table. All three appear in our first stage regression models, and hence their impact on attendance is incorporated into our expected home and away attendance numbers (the fitted values). The statistical significance of the parameters on all of these control variables indicates a direct impact on match outcomes independent of the impact on attendance levels at matches. These variables also matter over and above the impact of seasonal variation in the strengths of teams playing at home or away, as captures by the fixed effects. We find that actual attendance does not matter, but expected attendance does: an away win is less likely the larger is the expected home attendance, and is more likely the larger is the expected away attendance. Again, this suggests that, controlling for observed and unobserved variation in explanatory variables, home and away attendances, and hence crowd composition, does affect match outcomes, corroborating what Singleton et al. (2021) found in Egyptian football.

6 Conclusions

In this paper we document a novel attendance data set that can be used to extend our understanding of factors affecting sporting outcomes and their relationship with the nature of the crowd observing matches. Our data permit splitting the total crowd size into those that support the home team and the away team, revealing the extent to which the partisan nature of the crowd matters, and its composition. In some regression models, we document that fans are more strongly motivated by factors related to their own team, be it the home or away team. Results also indicate that the anticipated size of the home and away crowds attending a match have an impact on match outcomes over and above a range of other commonly used explanatory variables, and also explain unobserved variation in how well teams perform at home or away across seasons.

Little research on fan heterogeneity, in terms of the number of home and away fans attending sporting events, exists. Our promising preliminary results reveal the importance of fan heterogeneity in explaining match outcomes, as well as clear patterns in attendance decisions made by supporters of home and visiting teams. These results highlight the importance of additional research on the economic implications of fan heterogeneity. We hope that these results motivate additional research, and also motivate sports economists to collect and analyze data on home and visiting fan attendance in other settings.

References

V. Alfano. COVID-19 Diffusion Before Awareness: The Role of Football Match Attendance in Italy. *Journal of Sports Economics*, 23(5):503–523, 2022.

Anthony J Arnold. An industry in decline? The trend in football league gate receipts. *Service Industries Journal*, 11(2):179–188, 1991.

J. Borland and R. MacDonald. Demand for Sport. *Oxford Review of Economic Policy*, 19(4):478–502, Winter 2003.

A. Bryson, P. Dolton, J.J. Reade, D. Schreyer, and C. Singleton. Causal effects of an absent crowd on performances and refereeing decisions during Covid-19. *Economics Letters*, 2021.

Table 3: Results from estimating the away win regression model (6) when also decomposing the attendance effect into an expected and unexpected component. Fixed effects, where used, are for home and away team years. Clustering is at the home and away team year level.

	Dependent variable:					
	Away win outcome					
	(1)	(2)	(3)	(4)	(5)	(6)
Constant	0.528***	0.535***				
	(0.058)	(0.058)				
Log Home Attendance	−0.065***		0.067		−0.004	
	(0.008)		(0.043)		(0.047)	
Log Away Attendance	0.052***		−0.004		0.019	
	(0.006)		(0.013)		(0.013)	
Fitted Log Home Attendance		−0.068***		0.184**		−0.207*
		(0.008)		(0.087)		(0.117)
Fitted Log Away Attendance		0.055***		−0.027		0.044**
		(0.006)		(0.018)		(0.020)
Residual Log Home Attendance		0.031		0.032		0.031
		(0.051)		(0.049)		(0.049)
Residual Log Away Attendance		0.014		0.013		0.014
		(0.018)		(0.018)		(0.018)
Elo Prediction					−0.733	−0.441
					(0.719)	(0.718)
Balance					2.244**	2.169**
					(0.893)	(0.889)
Position (H)					−0.004*	−0.004*
					(0.002)	(0.002)
Position (A)					0.006***	0.006***
					(0.002)	(0.002)
Goals scored (H)					−0.006***	−0.005***
					(0.002)	(0.002)
Goals scored (A)					0.005***	0.005***
					(0.002)	(0.002)
Points from 6th (H)					−0.007***	−0.007***
					(0.003)	(0.003)
Points from 6th (A)					0.010***	0.009***
					(0.003)	(0.003)
Observations	6,873	6,873	6,873	6,873	6,873	6,873
R^2	0.016	0.017	0.430	0.431	0.444	0.445
Adjusted R^2	0.016	0.017	0.055	0.056	0.077	0.077
Residual Std. Error	0.454 (df = 6870)	0.453 (df = 6868)	0.444 (df = 4144)	0.444 (df = 4142)	0.439 (df = 4136)	0.439 (df = 4134)
Fixed effects	No	No	Yes	Yes	Yes	Yes
Note:						*p<0.1; **p<0.05; ***p<0.01

B. Buraimo and R. Simmons. A tale of two audiences: Spectators, television viewers and outcome uncertainty in Spanish football. *Journal of Economics and Business*, 61(4):326–338, 2009.

B. Buraimo and R. Simmons. 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, 2015.

B. Buraimo, D. Forrest, and R. Simmons. The 12th man?: Refereeing bias in English and German soccer. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 173(2):431–449, 2010.

J. Carbone, T. Corke, and F. Moisiadis. The Rugby League Prediction Model: Using an Elo-Based Approach to Predict the Outcome of National Rugby League (Nrl) Matches. *International Educational Scientific Research Journal*, 2(5):26–30, 2016.

A. Cardazzi, B.R. Humphreys, J.E. Ruseski, B. Soebbing, and N.S Watanabe. Professional Sporting Events Increase Seasonal Influenza Mortality in US Cities. Technical Report 3628649, SSRN, 2020.

D. Coates and B.R. Humphreys. Week-to-week attendance and competitive balance in the National Football League. *International Journal of Sport Finance*, 5(4):239, 2010.

D. Coates and B.R. Humphreys. Game attendance and outcome uncertainty in the National Hockey League. *Journal of Sports Economics*, 13(4):364–377, 2012.

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

C. Cueva. Animal Spirits in the Beautiful Game. Testing social pressure in professional football during the COVID-19 lockdown. Osf preprints, 2020.

P. Dawson and S. Dobson. The influence of social pressure and nationality on individual decisions: Evidence from the behaviour of referees. *Journal of Economic Psychology*, 31(2):181–191, 2010.

T. Dohmen and J. Sauermann. Referee Bias. *Journal of Economic Surveys*, 30(4):679–695, 2016.

A.E. Elo. *The rating of chessplayers, past and present*, volume 3. Batsford, London, 1978.

M. Endrich and T. Gesche. Home-Bias in Referee Decisions: Evidence from "Ghost Matches" during the COVID19-Pandemic. *Economics Letters*, (197):109621, 2020.

K. Fischer. Thinning out spectators: Did football matches contribute to the second COVID-19 wave in Germany? *German Economic Review*, 2022.

D. Forrest, R. Simmons, and B. Buraimo. Outcome uncertainty and the couch potato audience. *Scottish Journal of Political Economy*, 52(4):641–661, 2005.

P. Garcia-del Barrio and J.J. Reade. Does certainty on the winner diminish the interest in sport competitions? The case of Formula One. *Empirical Economics*, pages 1–21, 2021.

L. Garicano, I. Palacios-Huerta, and C. Prendergast. Favouritism Under Social Pressure. *Review of Economics and Statistics*, 87(2):208–216, 2005.

Celine Gimet and Sandra Montchaud. What drives European football clubs' stock returns and volatility? *International Journal of the Economics of Business*, 23(3):351–390, 2016.

K. Harb-Wu and A. Krumer. Choking under pressure in front of a supportive audience: Evidence from professional biathlon. *Journal of Economic Behavior & Organization*, 166:246 – 262, 2019. ISSN 0167-2681. doi: https://doi.org/10.1016/j.jebo.2019.09.001. URL http://www.sciencedirect.com/science/article/pii/S0167268119302756.

Stefan Késenne. The impact of salary caps in professional team sports. *Scottish Journal of Political Economy*, 47(4):422–430, 2000a.

Stefan Késenne. Revenue sharing and competitive balance in professional team sports. *Journal of Sports Economics*, 1(1):56–65, 2000b.

Stefan Késenne. The optimal competitive balance in a sports league? In P. Rodríguez, S. Késenne, and R. Koning, editors, *The Economics of Competitive Sports*. Edward Elgar Publishing, 2015.

N. Lowles and A. Nicholls. *Hooligans: AL of Britain's Football Gangs, Vol. 1*. Wrea Green: Milo Books, 2005.

D. McCarrick, M. Bilalic, N. Neave, and S. Wolfson. Home advantage during the COVID-19 pandemic: Analyses of European football leagues. *Psychology of Sport and Exercise*, 56:102013, 2021.

M. Olczak, J.J. Reade, and M. Yeo. Mass Outdoor Events and the Spread of an Airborne Virus: English Football and Covid-19. *Covid Economics*, September 2020.

K. Page and L. Page. Alone against the crowd: Individual differences in referees' ability to cope under pressure. *Journal of Economic Psychology*, 31(2):192 – 199, 2010. ISSN 0167-4870. The Economics and Psychology of Football.

D.A. Peel and D.A. Thomas. Outcome uncertainty and the demand for football: An analysis of match attendances in the English football league. *Scottish Journal of Political Economy*, 35(3):242–249, 1988.

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

P. Pettersson-Lidbom and M. Priks. Behavior under social pressure: Empty Italian stadiums and referee bias. *Economics Letters*, 108(2):212–214, August 2010.

J.J. Reade. A highly disaggregated look at competitive balance. In *Outcome Uncertainty in Sporting Events*. Edward Elgar Publishing, 2020.

J.J. Reade. Football attendance over the centuries. In *A Modern Guide to Sports Economics*, pages 21–43. Edward Elgar Publishing, 2021.

J.J. Reade, D. Schreyer, and C. Singleton. Eliminating supportive crowds reduces referee bias. *Economic Inquiry*, 2022.

D. Schreyer and P. Ansari. Stadium attendance demand research: a scoping review. *Journal of Sports Economics*, 23(6):749–788, 2022.

D. Schreyer, S.L. Schmidt, and B. Torgler. Football spectator no-show behavior. *Journal of Sports Economics*, 20(4):580–602, 2019.

Carl Singleton, James Reade, and Dominik Schreyer. A decade of violence and empty stadiums in Egypt: When does emotion from the terraces affect behaviour on the pitch? Discussion Paper 2021-21, University of Reading, 2021.

C. Stoecker, N.J. Sanders, and A. Barreca. Success Is something to sneeze at: Influenza mortality in cities that participate in the Super Bowl. *American Journal of Health Economics*, 2(1):125–143, 2016.

M. Sutter and M.G. Kocher. Favoritism of Agents - the Case of Referees' Home Bias. *Journal of Economic Psychology*, 25(4):461–469, August 2004.

S. Szymanski. Income inequality, competitive balance and the attractiveness of team sports: Some evidence and a natural experiment from English soccer. *Economic Journal*, 111(469):69–84, 2001.

S. Szymanski and S. Késenne. Competitive balance and gate revenue sharing in team sports. In S. Szymanski, editor, *The comparative economics of sport*, pages 229–243. Springer, 2010.

L. Vaughan Williams, C. Liu, and H. Gerrard. How well do Elo-based ratings predict professional tennis matches? Discussion Papers in Economics 2019/3, Nottingham Business School, 2019.

Pamela Wicker. The carbon footprint of active sport participants. *Sport Management Review*, 22(4): 513–526, 2019.