Department of Economics

Football Attendance Over the Centuries

By J. James Reade

Football Attendance Over the Centuries[*]

J. James Reade

Department of Economics

School of Politics, Economics and International Relations

University of Reading

March 10, 2020

Abstract

Attendance patterns at professional football matches in England and Wales over three centuries are analysed, with particular attention paid to the loyalty, or persistence, of spectators or fans at different football clubs over different periods of time. We find that about half of an attendance can be attributed to loyalty, and that patterns of loyalty have evolved considerably over time, and vary across regions within England and Wales, although there is little significant team-by-team difference. We also find that fans are not necessarily concerned about how balanced matches are expected to be, and that they are motivated by their own team's quality, and the geographic proximity of the visiting team.

1 Introduction

The demand for any good is of central importance, and the demand for professional football events is no different. What is rather unique when it comes to professional football is the length of time over which the numbers of people expressing a willingness to pay exist. For the English Football League, the oldest professional football league in the world and operating in England and Wales, attendance data has been recorded since the League began in 1888. Since 1888 139 different clubs have participated in the league, providing a vast amount of data over three centuries. Data on match outcomes back to 1888 also exists, allowing analysis of many of the determinants of attendance over a very long period of time.

In this paper, we analyse attendance at English Football League matches over three distinct centuries, employing panel time series methods in order to determine the extent to which patterns of attendance have changed over the years, and have varied from club to club. Of particular interest is what we might attempt to identify as habit persistence. Have football fans become more or less persistent in their attendance patterns over the years? Do fans differ between clubs, and across regions? These kinds of questions regarding the persistence of football spectator behaviour are naturally important for the financial viability of football.

We find variation across clubs, but more significantly around the regions of England and Wales, and over time, in the persistence of attendances. Attendance patterns have become markedly more persistent since the 1950s, and are particularly strong in the more traditional footballing parts of the country. The more we control for various other explanatory factors (competitive balance, refined measures of team quality, local interest), the less important is a change in league position for explaining attendance.

In Section 2 the relevant previous papers investigating demand for attendance are reviewed, in Section 3 the modelling methodology adopted is set out, in Section 4 our dataset and sources are introduced, in Section 5 results from the econometric estimations are presented, and Section 6 concludes.

2 Literature

Many studies considering the demand for attendance at sporting events have been conducted, and as such this review is bound to be only indicative of the range of areas covered, rather than exhaustive.

One of the first sports economics papers, Neale (1964), discusses various aspects of the demand for sport, from the *Louis-Schmelling effect* to illustrate the importance competitive balance, to the *fourth estate effect* and league standing effect, illustrating the interest generated in the press in sport, but also the impact of the league standing of sports teams as an indicator of their quality.

Hart et al. (1975) appears to be the first demand for attendance study, and considers Saturday afternoon attendance patterns at four top division English clubs between 1969 and 1972 (Leeds, Newcastle, Nottingham Forest and Southampton), and find that both geography, and the quality both teams matter significantly, although they attribute a significant proportion of variation in attendance to "a large irrational element in the psychology of football support".

Another early study of English Football League attendances was carried out by Bird (1982), at an historic low-point in attendances in the early 1980s. He found that demand was income inelastic, hence that at the time football was an inferior good. He also found that demand was price inelastic also.

Cairns (1987) considered structural changes in the Scottish league and their impact on attendance. He noted that changes towards smaller divisions in the early 1980s increased inequality in the distribution of attendances.

Peel and Thomas (1988) and Peel and Thomas (1992) both look at individual seasons within the English Football League (1981/82 and 1986/87 respectively), and in particular investigate the impact of competitive balance on attendance using bookmaker odds. In Peel and Thomas (1992) they introduce a measure of core support via lagged attendance numbers at particular clubs. In both cases they find that fans prefer to see their team win rather than a highly competitive match, controlling for other factors related to the quality of a given match. This is contrary to the widely-cited 'uncertainty of outcome hypothesis', a theory traced back as far as Neale (1964), that spectators prefer greater levels of uncertainty when consuming a sporting event.

Both Peel and Thomas (1996) and Szymanski (2001) consider 'natural experiments' of repeat fixtures between teams to evaluate attendance demand. Peel and Thomas (1996) consider Scottish leagues where teams play each other twice home and twice away, creating repeat fixtures. They find further evidence contrary to the 'uncertainty of outcome' hypothesis: fans prefer to see their team win rather than an exciting game. In England, the FA Cup often provides repeat fixtures when two teams in the same division are drawn against each other, and Szymanski (2001) use this to show the impact of the decreasing competitive balance in the FA Cup relative to the league on falling attendances — a more conventional 'uncertainty of outcome' finding.

Dobson and Goddard (1996) consider the demand for attendance at football matches across the regions of England and Wales, where the Football League operates, and takes advantage of time series methods to consider a long-run cointegrating relationship between attendance, ticket price, team performance, and a number of more economic variables. Their time dimension is seasons, rather than individual matches. They note that their price term can be removed from the long-run relationship, but that in the long term unemployment matters strongly. They suggest minimal differences across regions in terms of the loyalty of supporters, although they measure loyalty in terms of the short-run response of attendance to team performance, rather than simply by the persistence in attendance patterns as might be reflected by a lagged dependent variable such as that used by Peel and Thomas (1992).

Simmons (1996) makes use of time series methods within a panel of urban-based football clubs to conduct a more conventional demand analysis, finding evidence that more casual supporters are sensitive to price changes, and finding more generally that standard footballing variables such as league position do matter.

Peel and Thomas (1996) consider the situation where league competitions include repeat fixtures. This is true of the Scottish football league, where teams play each other four times, twice at each team's venue. Such repeat games control for many of the unmeasurable factors affecting demand. They consider one season of football, and all of the three divisions of Scottish professional football. Their findings suggest that supporters prefer to see their team win, rather than necessarily a highly uncertain match with a lower likelihood of victory.

Garcia and Rodríguez (2002) investigate the demand for attendance in Spanish league football, considering a panel of match-by-match data alongside extensive price data. The main objective of the paper is to estimate demand elasticities, in particular to investigate the common finding of inelastic price demand. They appear to confirm this finding, although they note that a range of estimates result from varying the specification of the regression model.

Forrest and Simmons (2002) consider in detail the extent to which the uncertainty surrounding outcome matters. They find that there is a non-linear relationship between uncertainty of outcome and attendance, namely that demand peaks when a home team is slightly more likely than usual to win a game, tailing off either side of this. They use betting prices to do this, and look at match-by-match data. They also note the interaction between competitive balance and home advantage, noting that by increasing equality across teams, the case of the home outsider facing a very strong visiting team would be lost, an occasion that they argue from their data does lead to greater attendance.

Forrest and Simmons (2006) conduct an analysis similar to that of Garcia and Rodríguez (2002) for English Football League teams around the turn of the twenty-first century, looking at match-by-match data for the seventy two teams that make up the Football League. They note that capacity constraints are much less binding in the Football League compared to the Premier League (a breakaway league of the top 20 teams that were previously in the Football League) enabling them to model using conventional least squares methods of estimation.

Dobson and Goddard (1995) would appear to be the study closest to this one, considering 94 teams over 67 seasons, finding a varied role for the price in the demand for attendance at different teams. They are unable to consider match-level data, however, as we do in this paper.

Price data, generally, is the most difficult data to collect when conducting demand studies; by and large, the best data any of the cited studies here have been able to use is gate receipts at the match level. This doesn't, however, control for the different prices that are often charged for admission to sporting events. For example, seating areas rather than standing areas, and concession prices for young and old

people. Reade (2019) and Nalbantis and Pawlowski (2018) are able to analyse data at individual ticket price level. Such data, naturally, is limited to particular clus, and particular eras, and hence the results may not necessarily be applicable more widely.

Baimbridge et al. (1996), Buraimo and Simmons (2008) and Buraimo and Simmons (2015) consider the distinction between spectators in the stadium, and watching on TV. Baimbridge et al. (1996) argue TV coverage reduces stadium attendance, but that the financial impact of this is positive, nonetheless. Buraimo and Simmons (2008) find that fans on TV prefer more balanced encounters, whereas Buraimo and Simmons (2015) find a stronger effect of quality of talent in events rather than how balanced they are.

Demand for attendance has been investigated for other popular British sports, such as cricket, despite cricket authorities releasing much less data on attendance than football. Schofield (1983) considered a new one-day league which emerged in England in the late 1960s, Hynds and Smith (1994) considered international test match cricket, Sacheti et al. (2014) also looked at international cricket, Sacheti et al. (2016) focussed on one-day international cricket, and Reade (2017) considered daily observations in the post-war era at the domestic level in England.

3 Methodology

This paper pursues an empirical methodology. The conventional approach to modelling demand in theory is absolutely fundamental to the basics of economics. Economic agents maximise their utility subject to a budget constraint, taking prices and income as given. An individual j maximises utility function $U_j = u(X_{j1}, p_1, \ldots, X_{jK}, p_K, M_j)$, where X_{jk} is the quantity of good k consumed by individual j, p_k is the price of good k, and M_j is individual j's income. Maximisation yields a set of demand functions $X_{jk}^* = f(p_1, \ldots, p_K, M_j)$ for each good. Conventionally, we then think about quantities consumed of goods, and prices, and infer characteristics of demand — elasticities of prices and of incomes.

In the context of demand for football attendance, individual level data do not exist. But data on the aggregate sums of people attending football matches exists back for, as we document in this paper, 140 years. As such, we model some att_{it}, the attendance for match i at time t, as $att_{it} = \sum_{j=1}^{J} X_{jit}$. We are considering counts of people paying to watch football matches in person at a football stadium.

We consider football clubs through time, and as such we have a panel time series setting, with both a large number of cross sections units (clubs, $N = 139$) and a large number of match-level observations ($T = 1413$ on average, maximum $T = 2482$, minumum $T = 11$). As such, the concerns regarding small N or T dimensions do not apply here, and we apply standard time series analysis methods to understanding the dynamics present, at a general level, in our dataset.

We run an autoregressive distributed lag (ADL) model for the log attendance at club i in match t, att_{it}, as a function of a vector of explanatory variables X_{it}. We introduce the autoregressive element, as Peel and Thomas (1992) and Dobson and Goddard (1996) do, in order to potentially capture persistence in attendance. The distributed lag element allows variables from previous matches to affect attendance on a given day, and enables the calculation of long-run effects.

We regress log attendance in order to interpret coefficients in terms of elasticities and semi-elasticities, where appropriate, and the transformation also helps with the differences in scale observed in attendance

in our sample. We specify P lags of both log attendance and the vector of explanatory variables:

$$att_{it} = \alpha_{0i} + \sum_{p=1}^{P} \alpha_p att_{i,t-p} + \sum_{p=0}^{P} \beta_p X_{i,t-p} + \varepsilon_{it}, \qquad \varepsilon_{it} \sim N(0, \sigma^2). \qquad (1)$$

We include a range of fixed effects, allowing the constant term can vary across units, hence α_{0i} has an i subscript. We check for how many are required. We allow $X_{i,t-p}$ to vary across p; for example X_{it} includes various fixed effects as specified later.

Persistence in attendance patterns is captured by the lagged dependent variable coefficients, hence α_p. In an AR(1) process, the α_1 coefficient would be interpreted as the percentage of a previous attendance that returns for the next home match; persistence in attendance patterns. With P lags the persistence is measured as $\sum_{p=1}^{P} \alpha_p$. The quality of inference into persistence depends on a number of assumptions, not least those placed on the error term, $\varepsilon_{it} \sim N(0, \sigma^2)$, holding true in the regression model. Particularly important will be the inclusion of fixed effects for years, and for clubs.

In the case where attendance is non-stationary, as would be reasonable to conclude, it is conventional to consider the transformed version of (1):[1]

$$\Delta att_{it} = \beta_0 \Delta X_{it} + (\alpha_1 - 1)\left[att_{i,t-1} - \kappa_{0i} - \kappa_k X_{i,t-1}\right] + \varepsilon_{it}. \qquad (2)$$

Here, we have assumed $P = 1$ for simplicity, and $\kappa_{0i} = \alpha_{0i}/(1 - \alpha_1)$ and $\kappa_1 = (\beta_0 + \beta_1)/(1 - \alpha_1)$. The expression $att_{i,t-1} - \kappa_{0i} - \kappa_1 X_{i,t-1}$ is the long-run relationship between attendance and the expanatory variables, and $\alpha_1 - 1$ represents the speed of adjustment back to equilibrium once disturbed away from it. The coefficient κ_1 can be expressed as the infinite sum of partial adjustments of attendance each successive time period to a change in X_{it}. As such, if we assume nothing else changes in the future, we can think about this as the long-run marginal effect.

We estimate (1) by ordinary least squares. Many attendance studies have estimated models of attendance using models for censored data, since many stadia, particularly in more recent times, are subject to capacity constraints, and as such we do not observe the true variation in demand. We appeal to the small proportion of our sample that is capacity constrained in defence of our estimation strategy, and also point out that we do not seek to interpret our results in terms of demand given our lack of price data.

The coefficients on the lagged dependent variables are fundamental to the analysis of the persistence of football attendance patterns. As such, we interact them with a range of controls in order to learn a little more about the nature of persistence. Football fans tend to be of the belief that a great deal can be determined simply by the knowledge of which team another person supports. What fans refer to as 'loyalty', we can consider to be a habit, or persistence in a pattern of attendance at a club level. Loyal supporters are those that persist with watching a team. Dobson and Goddard (1996) measure loyalty in terms of attendance response to performance measured by position, whereas Peel and Thomas (1992) measure 'core support', an obviously similar concept, by an autoregressive structure.

This loyalty notion suggests that, amongst other things, the team matters for persistence patterns, and so we test this by interacting lagged dependent variables with club fixed effects. Football supporters attach a slightly lesser degree of loyalty to their region, and indeed we might anticipate that with the regional spread of football over the three centuries of the Football League, that there are differences

---

[1]It is worth noting that the transformation remains valid statistically even if the data series are stationary in their levels.

in attendance patterns across regions. To investigate this we interact lagged dependent variables with regional fixed effects. Anecdotal evidence seems to suggest also that attendance patterns have differed over the decades, with more alternative leisure pursuits available in more recent times. To test this, we interact lagged dependent variables with a dummy variable for each decade.

The most salient explanatory variable for attendance is percieved to be league position. Although many other measures of team quality exist, and indeed are described in Section 4, league position is arguably the most simple, and most reflective. Rather than analysing various aspects of a visiting team, fans will make decisions based on whether that team is top of the league, is around about the top of the table, or worse. As a result, we focus on this in our analysis of our regressions, in addition to considering fan persistence patterns.

4 Data

We use data from the *www.11v11.com/* website, which describes itself as the 'Home of football statistics and history', and is maintained by the *Association of Football Statisticians*.

It contains results for all league matches that have taken place in England since 1888, as well as a number of other competitions (including European competition). Of the 224,932 matches recorded on the website as of December 2018, 182,728 have an attendance recorded for them. From the very first day of the Football League, on September 8 1888, it is recorded that 3,000 people watched Bolton Wanderers play Derby County, 12,000 people watched Everton play Accrington, and 2,500 people watched Wolverhampton Wanderers face Aston Villa.[2]

The one drawback of such a huge dataset of attendances is the unavailability of individual ticket prices for matches, meaning that we cannot estimate a demand function in the conventional economic sense.

In Figure 1 each of those 182,728 observations is plotted, and so each attendance is a dot in the diagram. The dataset covers the entire history of league football, and hence the early, pre-First World War, years, are a time of expansion from a low level in 1888. This would naturally have been a period of considerable expansion in the supply of observation areas at football clubs to cope with the revealed demand for football. This highlights the difficulty of identifying demand from supply in attendance data, even if some measure of ticket prices is available. Given that for us, price data is not available, we simply focus on explaining variation in attendance patterns, using as much data as we are able to make use of to explain observed patterns. After the Second World War, a steady decline is visible well into the 1980s, and the pick up since the 1980s appears quite broad based, albeit helped by particular growth in big attendances in a small number of constrained stadia, which is clearly visible on the plot.

When considering attendance patterns, we need to control for the level a match is at, as since 1893 the Football League has had multiple divisions, with teams separated on merit. From Figure 2 the average (calendar) yearly attendance is plotted, showing that higher level matches (black line) attract the highest attendances, with the difference being between 10 and 15 thousand. The dotted lines relate to the structure of the Football League between 1921 and 1958, as below the top two divisions the league was geographically split with a northern and southern division. After 1958 there was no geographical split, and simply four divisions based on merit. These lower two divisions, geographically or otherwise, have always had considerably lower attendances than the top two divisions, making clear the need to control for the level when explaining attendance.

---

[2]These early attendance figures must be rounded, but this is not the case in the vast majority of our dataset.
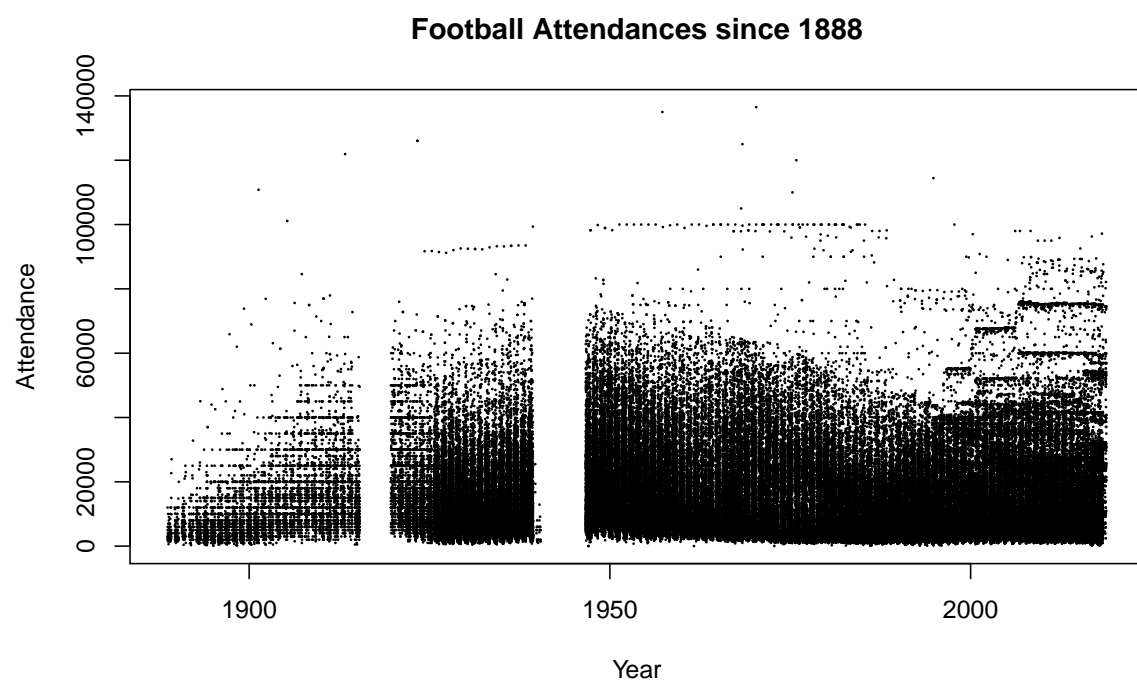
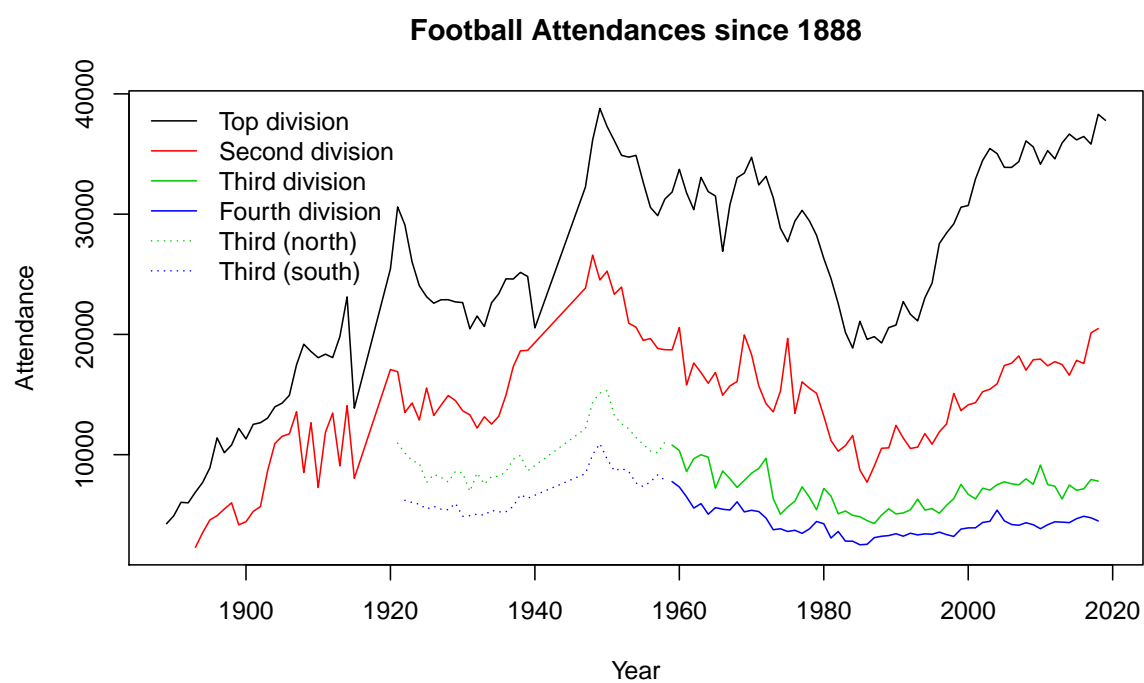Figure 1: All attendances through time in the Football League.

Figure 2: Average (over a calendar year) attendance by division through time in the Football League.
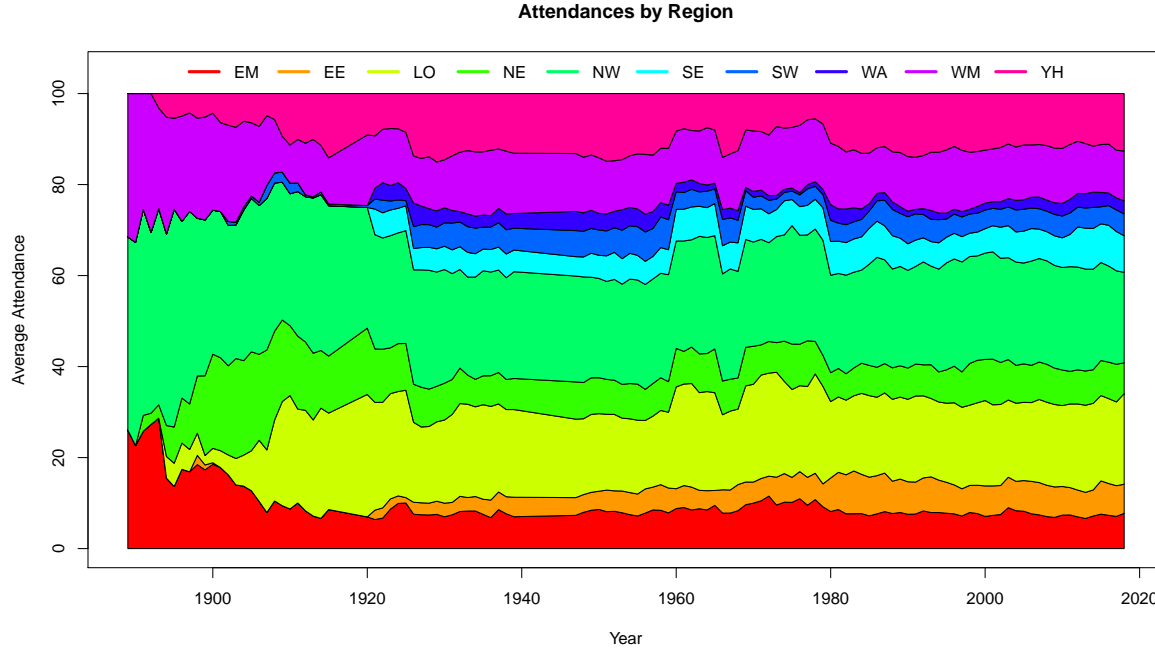
Attendances by Region

Figure 3: Average (over a calendar year) attendance by region through time in the Football League.

We also include team fixed effects, season fixed effects, and year fixed effects, to allow unobserved characteristics across each of these groupings to be reflected. League football began in the Midlands and North of England, and these areas remain traditionally strong in terms of attendance numbers; Figure 3 displays this regional variation, expressing the shares of total Football League attendance by region. Although by and large regional shares had settled into roughly their current levels by the end of the First World War, there is still variation through time, not least in the emergence of the East of England since that time. Because we include team fixed effects, we cannot include regional fixed effects, but these team fixed effects will pick up regional patterns such as those displayed in Figure 3.

There are 11 *NUTS1* regions throughout the entire UK, the broadest possible aggregation. These are areas such as the West Midlands, and the South East. Across the Football League, there are 36 *NUTS2* regions, and these correspond more broadly to counties, hence Lancashire, Merseyside and Kent. There are 90 *NUTS3* regions that have been represented by Football League clubs, and they represent towns, cities, and in the case of larger cities, parts of those cities. Around 14% of all matches are between teams in the same *NUTS1* region, almost 4% in the same *NUTS2* region, and just 1% in the same *NUTS2* region.

Because the *11v11* data includes the match result for every match, we make use of this to calculate a range of explanatory variables. Table 1 presents descriptive statistics for the explanatory variables that are used in the model. The average attendance over the entire sample is just under 14,000, with a standard deviation almost as large, giving some indication of the extent of the variation in the variable of interest. The ranges of other variables make sense; divisions have at most contained 24 teams in Football League history, while we measure form as the total points accrued in a team's previous six

9

Table 1: Descriptive statistics of data

Statistic	N	Mean	St. Dev.	Min	Max
Attendance	165,105	14,006.610	12,073.410	23	83,260
Log attendance	165,105	9.187	0.879	3.135	11.330
Position (home team)	196,507	12.066	6.551	1	24
Elo strength (home team)	196,507	1,074.815	141.688	637.372	1,579.160
Form (home team)	196,507	7.350	3.784	0	18
Goals scored (home team)	196,507	29.493	20.056	0	128
Position (away team)	196,507	11.662	6.558	1	24
Form (away team)	196,507	7.659	3.818	0	18
Goals scored (away team)	196,507	29.803	20.177	0	128
Elo Prediction	196,507	0.494	0.142	0.065	0.931
Match balance	196,507	−0.020	0.025	−0.189	0.000
Home team in contention	196,507	0.859	0.348	0	1
Points (Standard Deviation)	196,507	0.021	0.013	0.000	0.750
Dynasty	196,504	26.079	133.720	1	1,940
NUTS1 region derby	196,507	0.140	0.347	0	1
NUTS2 region derby	196,507	0.037	0.190	0	1
NUTS3 region derby	196,507	0.009	0.095	0	1
Days Since Last Match	196,368	19.237	95.688	0	16,544

matches, hence this can be at most 18.

We include measures of the quality of the home team. Football has a plethora of different measures of quality, and the value of the linear regression framework is that by including a number of them, we can determine which are most important. We include league position of the home team, yet this need not be the best measure of a team's quality within season, as it may be that teams have had an easier (harder) run of fixtures which means their position is somewhat inflated (deflated). To circumvent this problem we include *Elo strengths* of the home team. Elo strengths (Elo, 1978) are an alternative method for ranking teams that takes into account the strength of opponents that have been faced (which league standings do not do as they award points regardless of the opposition for particular match outcomes). If team A has true strength at time t of $R_{A,t}$ and team B has true strength at time t of $R_{B,t}$, then the expected score for team A against team B is:

$$E_A = \frac{1}{1 + 10^{(R_{B,t} - R_{A,t})/400}} = \frac{Q_A}{Q_A + Q_B}, \tag{3}$$

and the expected score for team B against team A is:

$$E_B = \frac{1}{1 + 10^{(R_{A,t} - R_{B,t})/400}} = \frac{Q_B}{Q_A + Q_B}, \tag{4}$$

where $E_A + E_B = 1$ and $Q_A = 10^{R_{A,t}/400}$ and $Q_B = 10^{R_{B,t}/400}$. Naturally, the true strengths of teams is unknown, and hence in practice one must choose a starting value for $R_{A,t}$ and allow it to be updated after each match. If the actual outcome of the match at time t for team A, $S_{A,t}$, differs from the expected outcome then that team's score needs updating; if $S_{A,t} = E_{A,t}$ then the existing strength for each team is accurate. While we could simply treat $S_{A,t}$ as binary, scaling the match outcome between

0 and 1 (as we do) allows information on the closeness of the match to be built into the updating process. Updating in the event of $S_{A,t} \neq E_{A,t}$ is done according to the formula:

$$R_{A,t+1} = R_{A,t} + K(S_{A,t} - E_{A,t}). \tag{5}$$

The factor K can be varied and is conventionally set at 32. The setting of K affects both the convergence of R_A to its true value and also the variation around that true value. A low K will lead to slow convergence, but only small variance around that value, whereas a higher K will lead to quicker convergence, but greater variance. Naturally, K can vary across different types of matches. For example, in a well-known international football variant of the Elo system, used in Reade and Akie (2013), FIFA World Cup Finals matches are weighted three times what international friendly matches are. We simply apply a weight of 40 to all matches.

When fans make decisions to attend or otherwise, they may do so based on simpler metrics than an Elo score. More likely, they reflect on a team's recent matches, what is usually referred to as their form. We include form for the home team and away teams, where form is measured as the total number of points accrued in a team's most recent six league matches. This form measure averages between 7 and 8 points in our dataset, with a standard deviation of around 3.8 points. Fans may also make a decision based on how many goals a team has scored up to that point in the season, as matches with more goals are generally considered to be more entertaining. We include this for both the home and away teams, and this variable averages between 29 and 30 goals, with a standard deviation of 20 goals.

How evenly matched two teams in a match are dictates the extent of the appeal of a match; the uncertainty of outcome hypothesis proposes that the greater the balance, the more interest there will be from spectators. We test that at three levels. First we consider how evenly matched the two teams are. This is done by taking a transformation of the Elo prediction, E_A, for a match. Noting that the most equal expected score would be $E_A = 0.5$, we thus consider the negative of the squared distance from 0.5 for E_A, hence $balance = -(E_A - 0.5)^2$. We take the negative so that the function is inverse-U shaped, achieving a maximum of zero for a match with $E_A = 0.5$. The lowest value this measure takes in our database is -0.208, and the average value is -0.021.

Then we consider season-level indicators: the standard deviation of points in the league (which on average is 0.021), and whether the home team remains in contention (that is, whether the team can reach the top three positions in the league given remaining points available). By this measure, almost 86% of matches involve the home team in contention. Finally we think about dynasties, where this is measured as the number of teams winning the League over the previous 25 years. On average this has been 3 teams, with a high of 21 and a low of 1 (which occurred only after the first ever season of the Football League.

Garcia and Rodríguez (2002) introduce the opportunity cost of attending matches, and we are also able to include non-football-related variables that reflect the appeal of a match. First, we add a variable for whether a match is between two teams from the same NUTS1, NUTS2 or NUTS3 region, reflecting the greater interest there usually is for matches between neighbouring teams; derby matches, as they are known. Fourteen percent of matches are between teams in the same NUTS1 region, 4% between teams in the same NUTS2 region, and 1% between teams of the same NUTS3 region.

We also add a variable for the number of days since the last home match to again capture opportunity cost. A shorter length of time between matches makes them more difficult to afford, potentially, and reflects a greater burden upon a potential attendee's time use. The mean value of this variable is 17.6 days, but it is heavily skewed by, for example, teams exiting the Football League and returning at a

later date. The median is 14 days between matches, the standard fortnight between home matches.

We don't include any macroeconomic variables in our estimation, such as GDP levels, growth, or unemployment rates. By including season fixed effects we crudely pick up the kind of impact of such variables over time.

5 Results

We present our results in steps, building up from a very basic model. This enables us to consider the impact of the additional variables we add. It is worth noting that such a basic-to-general approach runs counter to standard statistical principles; the most general model is most likely to satisfy the assumptions we place upon any statistical model, as expressed in (1). As such, while we do consider these small models, we do not emphasise their numerical values other than for the purposes of illustrating the bias induced by the omission of varialbes.

The most basic relationship to be posited, regarding attendance, is that it depends on league position. If a team is doing better, more people will attend. League position acts as a sufficient statistic as far as potential supporters are concerned. As such, we run a range of regressions of log attendance on league position, building from the most basic regression. We present these regressions as columns in Table 2, adding more and more variables as we read from left to right. The top panel are the regression coefficients, and where appropriate the fixed effects coefficients are suppressed. The panel beneath presents, where appropriate, the long-run solution coefficients as in (2). The bottom panels presents a small number of statistics evaluating the quality of the estimated model.

The leftmost regression simply regresses log attendance on league position, and suggests that a team gaining one league position can increase attendances by 3.6%, but that equally this variation in league position is only able to explain 7.4% of the variation in attendance over the years. Naturally, this is a rather meaningless number, as there is nothing in the model to distinguish Manchester United from Accrington Stanley. There are many ways to include this information, and we attempt to do so in a number of informative ways.

First, we include the lag of attendance, hence a club's most recent attendance as an explanatory variable. We also include dynamic information on league position also, via its lag. This regression is run over 14,000 fewer observations, reflecting that the data is somewhat patchy in places, and hence the most recent match may not have attendance recorded. Nonetheless, there are still 151,025 observations, and the lagged dependent coefficient is 0.86, suggesting that the attendance at any given match is about 86% of what it was at the previous match, plus a contribution from league position. The contribution of league position is smaller, from the contemporaneous variable at 1.4%, however this does not include the lagged information. Including this enables us to consider the long run impact of a change in position from the middle panel of the table; if a club could be one position higher, attendance would be about 3.9% higher, consistent with the previous estimate.

Adding dynamic information improved the model fit dramatically as would be expected, with the adjusted R^2 coefficient increasing to 0.8 from 0.07, and the standard error of the residuals falling by more than half. In the third column we add a second lag of the log attendance and league position variable to create a second-order autoregressive distributed lag model. The second lag is important, especially for attendance, suggesting that the persistence through time of football attendances is more complex than simply reflecting the most recent match.[3] Taking the sum of the two lagged dependent variables as the

---

[3]In addition, unreported though important, the second lag removes all autocorrelation from the residuals of the estimated

total persistence, the total effect is about the same (92%) but is distributed over the two most recent matches. About 7,000 observations are lost by adding an extra lag, although again model fit improves, and the impact of league position remains similar from the bottom panel, at 4.1%.

In the third, fourth and fifth columns, fixed effects are added. Figure 2 displays the importance of controlling for the division of the Football League that a club was competing in, and so we add in divisional fixed effects. These fixed effects have a further time dimension, as they relate to various league restructurings.[4] The persistence of attendance falls, with both the first and second lag coefficients markedly smaller and their total effect at about 78%. The long term impact of a change in position falls, too, from 4.1% to 3.1%, once we consider only changes within division. Naturally movements up divisions lead to larger crowds.

In the fourth column, we control for club-specific effects by adding team fixed effects. This addition further reduces the persistence of attendance to 69%, suggesting that with a simpler regression specification attendance patterns attributable to club-specific characteristics were wrongly being attributed to persistence. Adding in club-specific fixed effects reduces the long-run impact of a change in position to 2%.

In the final column, we allow for time-specific fixed effects; we include a fixed effect for each season of the Football League's existence. Consideration of both Figures 1 and 2 shows clear secular patterns throughout the Football League's existence, with initial growth, post-war peaks, a trough in the 1980s, and subsequent recovery. This addition reduces persistence yet further to 51%, makes little difference to the impact of position on attendance, and does affect the constant term, κ_0, by accounting for secular patterns in attendance levels over the centuries.

In sum then, adding these fixed effects, along with the club and division fixed effects, almost halves the effect of persistence. Much of what may have been attributable to habits and loyalty can better be explained by considering the division, the club, and the period of time that a given match took place in.

While unmeasurable club-specific and time-specific trends undoubtedly exist, it is important where possible to separate the unmeasurable from the measurable, and one the benefits of sport-related data is the ability to measure what is often considered to be unmeasurable in other economic contexts. As such, we add a number of control variables, although each of these variables is of interest in its own right, when thinking about determinants of attendances.

The first column of Table 3 presents the baseline model. The first six rows are the constant, lagged dependent variables and distributed lags of league position, as in Table 2. Persistence remains at about 50%, as it was in the final column of Table 2. In the next three rows are other measures related to the quality of the home team; their Elo strength, their form, and the number of goals the team has scored. The Elo strength variable has a mean of 1077.2, and hence is much larger than form, with an average value of 7.4, and as such the impact of Elo is proportionately greater, as well as being much more precisely measured. The impact of goals scored is negative, suggesting that once form and the intrinsic strength of the team as measured by position and Elo strength, the impact of goals is negative and probably indicative of a weaker side that also concedes more goals.

The following four rows provide a measure of the visiting team in a match, via their relative Elo strength (from the Elo prediction), their position, form and goals scored. The negative coefficient on

---

model.

[4]The second, third and fourth tiers were added gradually, with a geographical third tier after the First World War, which became simply a third and fourth tier in 1959. In 1992, the FA Premier League was created, a breakaway league from the Football League, and again in 2005 another renaming saw the division beneath the Premier League become the Championship.

Table 2: Table of regression output focussing on the role of dynamics and fixed effects.

| | Dependent variable: | | | | | |
| | lattendance | | | | | |
	(1)	(2)	(3)	(4)	(5)	(6)
Constant	9.625***	1.330***	0.756***	2.014***	2.562***	3.170***
	(0.004)	(0.011)	(0.011)	(0.017)	(0.055)	(0.079)
Log Attendance (lag)		0.864***	0.486***	0.414***	0.370***	0.280***
		(0.001)	(0.002)	(0.002)	(0.002)	(0.002)
Log Attendance (second lag)			0.438***	0.364***	0.319***	0.228***
			(0.002)	(0.002)	(0.002)	(0.002)
Position	−0.036***	−0.014***	−0.013***	−0.015***	−0.014***	−0.015***
	(0.0003)	(0.0003)	(0.0003)	(0.0003)	(0.0003)	(0.0003)
Position (lag)		0.009***	0.005***	0.004***	0.004***	0.002***
		(0.0003)	(0.0004)	(0.0003)	(0.0003)	(0.0003)
Position (second lag)			0.005***	0.004***	0.004***	0.002***
			(0.0003)	(0.0003)	(0.0003)	(0.0003)
$\widehat{\kappa_0}$		9.779	9.916	9.069	8.220	6.451
$\widehat{\kappa_1}$		-0.039	-0.041	-0.031	-0.020	-0.021
Observations	165,105	151,099	143,959	143,959	143,959	143,959
Adjusted R^2	0.074	0.802	0.851	0.861	0.867	0.879
Residual Std. Error	0.846 (df = 165103)	0.393 (df = 151095)	0.342 (df = 143953)	0.330 (df = 143941)	0.323 (df = 148812)	0.308 (df = 143693)
F Statistic	13,108.590*** (df = 1; 165103)	203,876.700*** (df = 3; 151095)	163,910.700*** (df = 5; 143953)	52,340.820*** (df = 17; 143941)	6,421.143*** (df = 146; 148812)	3,944.707*** (df = 265; 143693)
Division fixed effects	N	N	N	Y	Y	Y
Team fixed effects	N	N	N	N	Y	Y
Season fixed effects	N	N	N	N	N	Y

Note: *p<0.1; **p<0.05; ***p<0.01

14

the visiting team position implies that the stronger is a visiting team, the larger is the attendance; each position higher increasing attendance by 0.7%. Form and goals scored by the visiting side prior to the match have negative coefficients also, indicating that spectators would like to watch the home team play out-of-form visiting sides that don't score too many goals, after controlling for their league position. The Elo prediction has a negative coefficient, which suggests that the more likely the home team is to win (the larger is the prediction variable), the less likely are fans to attend (after controlling for all other variables included).

The next four rows relate to measures of competitive balance. There is no impact of dynasties, as measured by the number of different winners of the League over the previous 25 years. The standard deviation of points in the league table, which is higher in the case of less competitive leagues, has the anticipated effect: a less competitive league lowers attendances at matches within it. The third level of uncertainty is match-level, and we capture this by both the balance implied by the Elo ranking, and a dummy variable for whether the home team remains in contention to finish in the league's top three positions. If the home team is still in contention, attendances are higher by between 2.5 and 3%, while the more balanced is a match, the lower is attendance. This is a pattern noted elsewhere, not least in Buraimo and Simmons (2008) where fans in the stadium are found to prefer their team to win, whereas fans watching on TV would prefer a more evenly matched event.

The following four rows are non-football-related variables; we have dummies for NUTS1, NUTS2 and NUTS3 regions specific matches, or derbies as they are colloquially known. A game between clubs in the same NUTS1 region has about a 11% greater attendance. Games between clubs in the same NUTS2 geographic areas have a further 18% larger attendances. Matches between teams within NUTS3 areas, have another 14% more spectators in attendance.

Finally, each extra day between matches increases the attendance by 0.02%, reflecting that matches where there is only a few days inbetween will attract fewer than if there has been a longer period of time since the last match. This indirectly reflects the budget constraints of supporters.

Adding these control variables reduces the long-run impact of position on league position further to just 1.1%.

The second, third and fourth columns interact the lagged dependent variable with the fixed effects for secular trends (inserting only decade indicators rather than a fixed effect for each year), regional effects, and club effects. The bottom two rows display which fixed effects are interacted, and report an F test for the significance of those fixed effect interactions. In all three cases, the null that they are zero is rejected.

From the second column, persistence has changed over the decades. The variation is perhaps best expressed graphically, and the sum of the autoregressive parameters related to each decade is presented in the top panel of Figure 4. Recorded persistence increased distinctly between the 1950s and the 1970s, which corresponds to the period of the most precipitous decline in attendances from Figure 1. This greater persistence in attendance patterns has continued at this higher level of 0.55 since the 1970s.

A staple of football supporter interactions is the notion that some supporters are more loyal than others. Loyalty is the idea that fans are persistent in attending regardless of how the team is performing. In the third column we consider whether there are regional disparities in persistence, and hence 'loyalty'. We plot the differences across regions in the middle panel of Figure 4. Supporters in Wales are the most 'loyal', followed by those in the East Midlands, Yorkshire and the Humber and the North West. Persistence is lowest in London, the South West and East of England.

The final column interacts club fixed effects with teams. This is the bottom panel in Figure 4, but

Table 3: Table of regression results interacting persistence with various other factors. Fixed effects for seasons, clubs, regions and divisions included.

|	*Dependent variable:*			
	Log Attendance			
	(1)	(2)	(3)	(4)
Constant	4.854***	6.595***	11.183***	8.233***
	(0.210)	(0.786)	(1.285)	(0.678)
Log Attendance (first lag)	0.273***	−0.053	0.293***	0.242***
	(0.002)	(0.069)	(0.006)	(0.091)
Log Attendance (second lag)	0.224***	0.192***	0.222***	0.434***
	(0.002)	(0.065)	(0.006)	(0.095)
Position (home team)	−0.011***	−0.010***	−0.011***	−0.010***
	(0.0003)	(0.0003)	(0.0003)	(0.0003)
Position (home team, first lag)	0.003***	0.003***	0.003***	0.003***
	(0.0003)	(0.0003)	(0.0003)	(0.0003)
Position (home team, second lag)	0.002***	0.002***	0.002***	0.002***
	(0.0003)	(0.0003)	(0.0003)	(0.0003)
Elo Strength (home team)	0.001***	0.001***	0.001***	0.001***
	(0.00002)	(0.00002)	(0.00002)	(0.00002)
Goals Scored (home team)	−0.001***	−0.001***	−0.001***	−0.001***
	(0.0001)	(0.0001)	(0.0001)	(0.0001)
Form (home team)	0.005***	0.005***	0.005***	0.005***
	(0.0003)	(0.0003)	(0.0003)	(0.0003)
Elo Prediction	−0.498***	−0.524***	−0.498***	−0.500***
	(0.013)	(0.013)	(0.013)	(0.013)
Position (away team)	−0.006***	−0.006***	−0.006***	−0.006***
	(0.0002)	(0.0002)	(0.0002)	(0.0002)
Form (away team)	−0.006***	−0.005***	−0.006***	−0.006***
	(0.0003)	(0.0003)	(0.0003)	(0.0003)
Goals Scored (away team)	−0.0004***	−0.001***	−0.0004***	−0.0004***
	(0.0001)	(0.0001)	(0.0001)	(0.0001)
Dynasties	0.00000	0.00000	0.00000	0.00001
	(0.00001)	(0.00001)	(0.00001)	(0.00001)
Match Balance	−0.931***	−0.893***	−0.931***	−0.953***
	(0.031)	(0.031)	(0.031)	(0.031)
In Contention (home team)	0.024***	0.029***	0.024***	0.024***
	(0.003)	(0.003)	(0.003)	(0.003)
League Points Standard Deviation	−0.279***	−0.365***	−0.273***	−0.243***
	(0.072)	(0.071)	(0.072)	(0.071)
NUTS1 Derby	0.112***	0.111***	0.113***	0.112***
	(0.003)	(0.003)	(0.003)	(0.003)
NUTS2 Derby	0.176***	0.175***	0.176***	0.176***
	(0.005)	(0.005)	(0.005)	(0.005)
NUTS3 Derby	0.136***	0.135***	0.136***	0.135***
	(0.009)	(0.009)	(0.009)	(0.009)
Days Since Last Home Game	0.0002***	0.0001***	0.0002***	0.0001***
	(0.00001)	(0.00001)	(0.00001)	(0.00001)
$\widehat{\kappa}_0$	5.239	6.125	5.504	2.942
$\widehat{\kappa}_1$	-0.011	-0.006	-0.012	-0.017
Observations	143,956	143,956	143,956	143,956
Adjusted R^2	0.895	0.897	0.895	0.898
Residual Std. Error	0.286 (df = 143677)	0.284 (df = 143639)	0.286 (df = 143659)	0.283 (df = 143425)
Persistence with	N/A	Decade	Region	Club
F Test	N/A	64.884*** (df = 38; 143639)	23.671*** (df = 18; 143659)	14.26*** (df = 252; 143425)
Note:				*p<0.1; **p<0.05; ***p<0.01
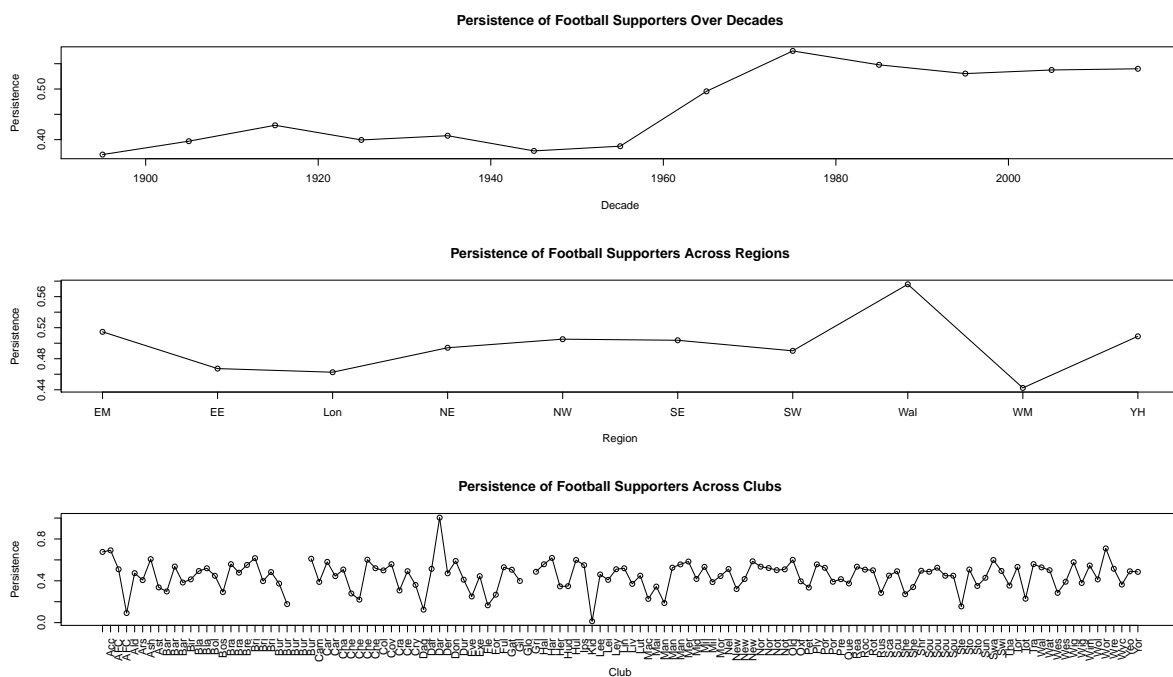
Figure 4: Variation in the persistence of attendance patterns across decades and across regions in the Football League.

with 130 clubs it is not possible to distinguish each team. A number of the outliers are teams with very few observations, relatively (those with short stays in the Football League). The significant mass of observations are within a range between 0.4 and 0.6, and as such as statistically indistinguishable from each other.[5]

6 Conclusions

Attendance patterns across three centuries are considered in this paper, using almost 200,000 match-level observations on over one hundred Football League clubs. In the absence of price data, focus is instead placed on the persistence of attendance patterns. Adequate controlling for variation attributable to the quality of the teams involved in the match, as well as other factors such as those related to competitive balance, we find a smaller, but nonetheless significant persistence effect that has varied over time, and also varies across regions, and teams.

References

M. Baimbridge, S. Cameron, and P.M. Dawson. Satellite Television and the Demand for Football: a Whole New Ball Game? *Scottish Journal of Political Economy*, 43:317–333, 1996.

---

[5] According to this measure, the following teams have the most loyal fanbases: Accrington Stanley, Hartlepool, Brighton, Bury, Chester, and Oldham, and the following teams have the least loyal: Dagenham and Redbridge, Stevenage, Fleetwood, Burton, Manchester City, Cheltenham, Macclesfield Town, Tottenham Hotspur and Everton.

P.J.W.N. Bird. The demand for league football. *Applied economics*, 14(6):637–649, 1982.

B. Buraimo and R. Simmons. Do sports fans really value uncertainty of outcome? evidence from the english premier league. *International Journal of Sport Finance*, 3(3):146, 2008.

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.

J.A. Cairns. Evaluating changes in league structure: the reorganization of the Scottish Football League. *Applied Economics*, 19(2):259–275, 1987.

S.M. Dobson and J.A. Goddard. The demand for professional league football in england and wales, 1925-92. *The Statistician*, pages 259–277, 1995.

S.M. Dobson and J.A. Goddard. The demand for football in the regions of England and Wales. *Regional studies*, 30(5):443–453, 1996.

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

D. Forrest and R. Simmons. New Issues in Attendance Demand: The case of the English Football League. *Journal of Sports Economics*, 7(3):247–263, 2006.

David Forrest and Robert Simmons. 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, 2002.

J. Garcia and P. Rodríguez. The determinants of football match attendance revisited empirical evidence from the spanish football league. *Journal of Sports Economics*, 3(1):18–38, 2002.

R.A. Hart, J. Hutton, and T. Sharot. A Statistical Analysis of Association Football Attendances. *Applied Statistics*, 24(1):17–27, 1975.

M. Hynds and I. Smith. The Demand for Test Match Cricket. *Applied Economics Letters*, 1(7):103–106, 1994.

G. Nalbantis and T Pawlowski. Game Uncertainty and the Demand for Quality Seating. Conference presentation, XIII Gijon Conference on Sports Economics, May 2018.

W. C. Neale. The Peculiar Economics of Professional Sports. *Quarterly Journal of Economics*, 78(1):1–14, 1964.

D. Peel and D. 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. Peel and D. Thomas. Attendance demand: an investigation of repeat fixtures. *Applied Economics Letters*, 3(6):391–394, 1996.

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

J.J. Reade. Daily Demand: Leicestershire Cricket in the post-war era. Discussion paper, University of Reading Department of Economics, 2017.

J.J. Reade. Paying on the Gate: Turnstile by turnstile analysis of attendance demand at football. Discussion paper, University of Reading, January 2019.

J.J. Reade and S. Akie. Using Forecasting to Detect Corruption in International Football. Working Papers 2013-005, The George Washington University, Department of Economics, Research Program on Forecasting, May 2013. URL http://ideas.repec.org/p/gwc/wpaper/2013-005.html.

A. Sacheti, I. Gregory-Smith, and D. Paton. Uncertainty of Outcome or Strengths of Teams: An Economic Analysis of Attendance Demand for International Cricket. *Applied Economics*, 46(17): 2034–2046, 2014.

A. Sacheti, D. Paton, and I. Gregory-Smith. An Economic Analysis of Attendance Demand for One Day International Cricket. *Economic Record*, 92(296):121–136, 2016.

J.A. Schofield. The Demand for Cricket: The Case of the John Player League. *Applied Economics*, 15 (3):283–296, 1983.

R. Simmons. The Demand for English League Football: A Club-Level Analysis. *Applied Economics*, 28 (2):139–155, February 1996.

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.