Sports Economic Review 2 (2023) 100012

Contents lists available at ScienceDirect

Sports Economic Review

journal homepage: www.journals.elsevier.com/sports-economic-review

Ghost games and ex-post viewing preferences for the English Premier League: Evidence from YouTube highlights☆

David Butler, Robert Butler *

Department of Economics, University College Cork, Cork, Ireland

A R T I C L E I N F O

JEL classification:
Z20

Keywords:
Ghost games
COVID-19
Demand
Streaming
YouTube
Broadcasting

A B S T R A C T

We ask if the move to play elite football (soccer) matches without spectators during the COVID-19 pandemic impacted online viewership. Using a new dataset from YouTube, we depart from the traditional ex ante approach to modelling football demand and investigate streaming preferences for known results in football. Our data is collected in real-time and considers English Premier League matches from 2019 to 2021 played both in front of crowds and behind closed doors. The results indicate increased viewership for matches played without crowds. The findings also allow a deeper understanding of direct demand for football as we identify motivators, including different pre-game and in-game characteristics as well as scheduling effects, that make fans curious to view content after the fact. The research adds to our understanding of the effects of the pandemic on fan viewer behaviour and speaks to broadcasting firms developing digital extension strategies.

1. Introduction

Since the COVID-19 pandemic, a growing literature has evaluated the impact of behind closed door (BCD) sporting events. The policies that prevented fans from attending live sport offered a natural experiment to evaluate a suite of topics important to the decision-making (see Leitner et al. (2022) for a recent review). Many effects of absent crowds have been investigated, most notably, the topic of home advantage (Cross & Uhrig, 2023; De Angelis & Reade, 2022; Fischer & Haucap, 2021; Gong, 2022; Losak & Sabel, 2021). Other topics of interest have been odds pricing (Winkelmann et al., 2021), attendance demand (Reade et al., 2021), referee bias (Bryson et al., 2021; Reade et al., 2022; Wolaver & Magee, 2022) and performance (Caselli et al., 2021; Farnell, 2023; Ferraresi & Gucciardi, 2021). Despite the recent growth in this literature, little is known about how fan broadcasting preferences reacted to BCD matches. We take advantage of the exogenous variation caused by COVID-19 to address an increasingly important aspect of broadcasting - streaming demand. We do this by capitalising on the public disclosure of YouTube views for English Premier League (EPL) matches and model highlight reel views for the 2019/20 and 2020/21 seasons that switched between in-person attendances and BCD.

To our knowledge, this research represents the first empirical study to consider broadcast preferences for football in light of the COVID-19 pandemic and is the first to model online views of football fans via a digital third-party platform. Demand via streaming is a neglected area of the literature. Our YouTube viewership data is advantageous as it is sourced from a reputable online channel (Sky Sports) and represents 'on-demand' consumption that incorporates mobile media streaming. Furthermore, our dataset is exclusive as viewership is measured in real-time; we follow the EPL game schedule on a rolling basis and extract YouTube views prior to the start of a new game-week both before and after the introduction of BCD matches.[1]

We also make an auxiliary contribution to the economics of sports broadcasting literature. Beginning with Forrest et al. (2005) studies which model EPL audience size offer an ex ante assessment of demand. Given the approach, this literature typically controls for pre-match characteristics, motivating fans to view linear (television) broadcasts. Recent examples of EPL broadcasting studies in this tradition include Buraimo and Simmons (2008), Scelles (2017), Cox (2018) and Buraimo et al. (2021). The demand modelling in these works' is premised on

☆ This research did not receive any grants from funding agencies in the public, commercial or not-for-profit sectors.
* Corresponding author.
E-mail addresses: david.butler@ucc.ie (D. Butler), r.butler@ucc.ie (R. Butler).

[1] These observations from YouTube cannot be retrieved without access to private channel analytics (available only to YouTube/Sky UK Limited). It is our understanding that YouTube analytics API's can only extract historical daily view data for channels and not individual videos. Given the nature of our data we are happy to make the streaming measures available on request.

https://doi.org/10.1016/j.serev.2023.100012
Received 7 February 2023; Received in revised form 24 April 2023; Accepted 25 April 2023
Available online 26 April 2023

unknown match outcomes. To date, there is little knowledge of what motivates fans to watch content ex post. As we can also offer an investigation of why fans demand content after the fact, our study has practical relevance for broadcasters.

The paper continues as follows. The next section contextualises the study from a broadcasting perspective. Section 3 describes our data, variables and empirical model. The results are then presented and discussed in section 4. The paper concludes with a brief discussion in addition to practical recommendations, limitation of the study and avenues for future research.

2. Background

While our primary focus is the impact of ghost games, it is important to establish a context for consumer demand insofar as it concerns viewing repeat content. Despite a long history of broadcasting demand, understanding the determinants of fans motivation to watch after-the-fact is yet to receive empirical attention. In the United Kingdom (UK), *Match of the Day* began in August 1964, originally broadcasting live matches on BBC2 (Smith, 2004). This programme still airs matchday highlights every Saturday night. This programme has now also spawned counterpart shows including *Match of the Day 2* (Sunday night) and a U.S. version aired on NBCSN.

At least historically, fan demand to view this content was impacted by broadcasting rules. For example, matches played on Saturday 3pm cannot be broadcast live on TV in England due to the "blackout" rule - a complete ban on the live broadcasting of any football match in the UK between 2.45pm and 5.15pm on Saturday.[2] The origins of this rule can be found in the early 1960s. Chairman of Burnley Football Club, Bob Lord, was concerned live televised football would negatively impact match attendance and successfully lobbied Football League executives to prevent the broadcasting of live games on Saturday afternoon throughout the UK (Jones, 2019).

Today, domestic media rights for the EPL have generally become increasingly more valuable (Butler & Massey, 2019; Feuillet et al., 2019; Scelles et al., 2020). While contracts to broadcast rights for known results pale in comparison to live broadcasting deals, the UK's state broadcaster - the BBC - regularly renew contracts with the EPL for ex-post broadcasting content. For example, the BBC recently paid an increased fee of £211.5m (from £204m) to broadcast EPL highlights through a suite of its highlights shows (BBC, 2018). These broadcasts remain resilient in the face of online competition and attract a non-trivial level of viewership – *Match of the Day* and *Match of the Day 2* attract a joint audience greater than seven million people each weekend (BBC, 2018).

While the BBC in the UK is a public corporation, online providers have strategic economic motivations for hosting this content. Fundamentally, providing free content ex-post represents an advertising strategy for private broadcasters; offering samples (e.g. free clips, highlight reels, post-match interviews) acts as a conduit to transition free users to premium subscriptions. Added financial benefits include access to advertising revenues, the potential to leverage network externalities and scope to grow brand equity through more online interactions.

3. Empirical framework

In August 2019 Sky UK Limited widened their content options for non-subscribing consumers by hosting highlights on their official *Sky Sports Football* YouTube channel. This move presented a new data source which is the basis of this study. Our pooled data set covers match-level data for 751 EPL matches, over the 2019/2020 and 2020/21 seasons, since the new Sky Sports policy was introduced. BCD matches constitute 57% of observations. This presents a natural experiment to consider the relationship between online views and BCD matches. The first BCD observation is recorded on the resumption of the Premier League in June 2020. BCD matches lasted until August 2021, when full crowds were allowed to return to stadia. This figure excludes matches that were allowed a restricted number of attendees. This occurred for a select number of clubs in December 2020. We later check for any impact of restricted attendances by treating these as BCD matches.

Our data is compiled from the following websites between August 2019 and May 2021: YouTube.com, Premierleague.com, Skysports.com, Oddsportal.com, Understat.com and Whoscored.com. Nine matches (Game-week 34) of the 2020/21 season were not uploaded to YouTube, so the full two seasons (760 matches) were not available.

3.1. YouTube views

As a broadcast medium, YouTube data is different to television audience figures that are typically modelled in ex-ante studies. Our data represents consumption through all digital devices capable of streaming content and does not exclude out-of-home viewing. Also, exact viewer data-points can be extracted from YouTube. This differs from rounded television audience figures that characteristically represent averages over a broadcast and can often traverse an entire programme, including pre/post-match analysis (Buraimo et al., 2021).

We record views in real-time for each match approximately 24 h before kick-off on the subsequent game-week. For example, if two game-weeks consecutively began on Saturday at 3pm GMT, the data consist of views between the time of upload (~5.15pm on a Saturday) to the following Friday when measurement takes place (between 3pm and 5pm). An equivalent strategy is followed for game-weeks beginning during the week. A weekly strategy is pursued as we reason that this is the product's 'best-before-date' and represents the club's most recent EPL performance. For each weekly set of observations, the maximum imposed time period for viewership is six days, although the EPL may not return because of international breaks.[3] Over the two EPL seasons we observed a non-trivial number of weekly views - by May 2021, EPL matches had accumulated approximately 450 million views.

We also measure view counts at a second constant point in real-time. Views are extracted for the complete 2019/20 season on 2nd August 2020, one week after the conclusion of the 2019/20 EPL season in July 2020 and on the 30th of May 2021, one week after the conclusion of the 2020/21 EPL season.[4] This end-of-season view data allows us to consider any bias resulting from delays in uploads or minor time variations related to extracting the weekly data in real-time. It also allows us to control for scheduling differences between match kick-off times. We scale this end-of-season measure using the log of days since a reel was uploaded, allowing us to account for the time available to potentially view the content. Both YouTube streaming measures are displayed in Fig. 1.

We have confidence in the accuracy of YouTube view count. Presently, the count algorithm ensures that views only reflect human clicks and videos (greater than 301 views) are regularly validated by the platform. It is also our understanding that a view is valid after a human intentionally engages the 'play' button via one device and has a total watch time of a minimum of 30 s or interacts with an advert. This count function is critical to YouTube's monetization model, as higher viewed videos are more likely to attract advertisers. YouTube can then make inferences as to which videos to monetize based on popularity. Additionally, this count provides a direct signal to content creators regarding the popularity of their content and their potential to monetize the platform. Thus, YouTube face a strong incentive to ensure reliable and accurate tallies.

[2] The blackout rule was not applied for BCD games during Covid-19.

[3] Our analysis doesn't extend to each reel's metadata (i.e. tags and descriptors).

[4] The EPL was postponed in March 2020 due to the Covid-19 pandemic and returned to complete the 2019/20 behind closed doors in June 2020. The 2019/20 did not finish until July 2020.

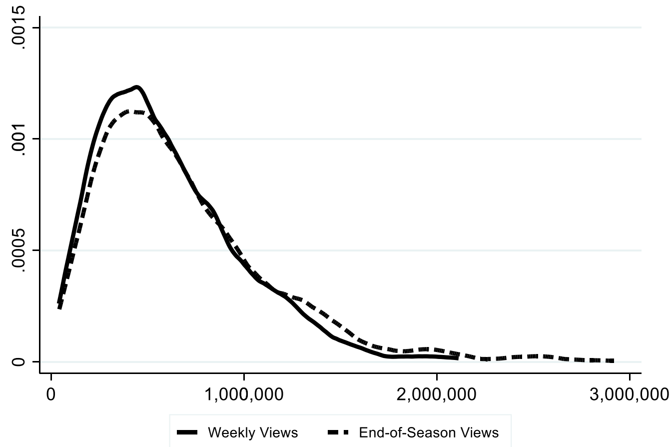

Fig. 1. Kernel density estimate of weekly and end-of-season streaming views.

3.2. Measures

Our primary focus concerns the impact of BCD matches – inclusive of reduced attendances – on YouTube streaming. This is simply captured by an indicator variable. However, as we are considering demand when outcomes are known, we must control for a variety of match-level characteristics that could determine views ex-post and outside of our key variable of interest.

We measure 'shock' by utilising exact score-line probabilities available on oddsportal.com. While these probabilities are unadjusted, they quantify audience surprise at an outcome continuously and controls for the marginal goal difference in a match result. We expect fans to show less interest in matches where the score line result is more predictable a priori. Shock results generally constitute both unlikely outcomes and high-scoring matches. It does not require a pre-match favourite to lose.[5]

We employ the sum of expected goals (xG from Understat.com), as opposed to the number of goals scored in a match, to capture match quality. xG is favoured over goals as it is more informative in judging performances and offers a richer measure for the quality of chances created (Kharrat et al., 2020). We also appeal to many additional in-play measures such as hat-tricks (scoring three times in a game), serious foul play (red cards), goals from star players, video-assistant referee controversies (VAR), serious injuries, impressive saves from goalkeepers, bad mistakes (howlers), exceptional goals, penalties and near misses.

Many of these motivators for viewing are less conspicuous but could be accurately recorded through textual analysis of weekly commentary timelines available from Sky Sports.com. This source provides specific icons to indicate injuries, shots of the crossbar/post and exceptional goalkeeping. The selection of our scheduling variables is informed by consumer modelling theory insofar as it concerns direct demand for live sport (Borland & MacDonald, 2003; Szymanski, 2003).

Table 1 describes the additional variables of theoretical importance to the modelling. The suite of in-play variables we identify are, for the most part, dummy indictors resulting from a mixture of basic intuitions of what in-play events fans likely prefer (e.g., exceptional goals, impressive saves, serious foul play, referee mistakes, etc.) and empirical investigations of specific fan preferences such as stardom (Buraimo & Simmons, 2015; Yang & Shi, 2011; Jewell, 2017). In addition, we

[5] For example, a favourite may successfully win a match. However, if they win by a high goal margin this may still represent a shock for fans. Twice in our dataset two pre-match favourites won 9-0. The match with the lowest shock value was a Wolverhampton Wanderers 1-0 victory over AFC Bournemouth in the 2019/20 season (19.16% probability of 1-0). The match with the highest shock value was Leicester City's 9-0 away win over Southampton in the 2019/20 season (probability of 0–9 = 0.25%).

Table 1
Variables.

Variables	Source	Description
BCD	Premierleague.com	1 = if a match was played behind closed doors.
SHOCK	Oddsportal.com	Pre-match score line odds probability of observed outcome.
EXPECTED GOALS (xG)	Understat.com	Match level metric that attributes the probability of a shot resulting in a goal.
PRE-MATCH		
Big6	Premierleague.com	1 = if two of the 'Big6' compete (Arsenal, Chelsea, Liverpool, Manchester City, Manchester United, & Tottenham Hotspur).
Derby	Skysports.com, Premierleague.com	1 = if the match was a derby.
NewManager	Skysports.com, Premierleague.com	1 = if the match featured a new manager (inclusive of caretakers).
Critical Fixture	Multiple Media Sources; Skysports.com, BBC.com, Premierleague.com	1 = if the match is important to a title win, Champions League/Europa Cup qualification or relegation.
IN-PLAY EVENTS		
StarScore	FIFA EA Sports, Skysports.com	1 = if a star player scores.
Hat-Trick	Skysports.com	1 = if a player scores three goals.
VARCon	Skysports.com, BBC.com	1 = if the match is affected by a controversy related to the decision of the video assistant referee.
RedCard	Skysports.com	1 = if a player or manager is sent off.
Save	Skysports.com	1 = if a goalkeeper makes an impressive save.
Injury	Skysports.com	1 = if a player incurs a serious injury.
Howler	Skysports.com, Whoscored.com	1 = if a player makes a critical error.
GOTM	Premierleague.com	1 = if a match goal is in the official EPL Goal of the Month poll.
Penalty	Skysports.com	1 = if a penalty is awarded.
NearMiss	Skysports.com, Whoscored.com	The total number of shots off the crossbar or goalpost and goal-line clearances that do not lead directly to a goal.
SCHEDULING		
Game-week	Premierleague.com	Game-week number.
Game-week^2	Premierleague.com	Game-week number squared.
Holiday	Defined by calendar	1 = Christmas & New Year Holidays.
Friday	Defined by calendar	1 = if a match was on a Friday night.
Saturday	Defined by calendar	1 = if a match was on a Saturday.
Sunday	Defined by calendar	1 = if a match was on a Sunday.
Monday	Defined by calendar	1 = if a match was on a Monday night.
Midweek	Defined by calendar	1 = if a match was on a Tuesday to Thursday.
SubTV	Premierleague.com	1 = if a subscription was required to view a match live (Sky Sports, BT Sports, Amazon Prime).
BoxOffice	Premierleague.com	1 = if a match was broadcast with an additional once-off PPV fee.
FTATV		1 = if a match was broadcast on free-to-air television.

Table 2
Event definitions.

Variable	Definition
VARCon	Contentious VAR decisions leading to a significant time delay and extensive media exposure.
Howler	An embarrassing miss/error when a player faces a clear goal-scoring opportunity or an obvious mistake leading to a goal.
StarScore	We defined star players using Electronic Arts (EA) FIFA rankings (seasonal release of the simulation) scoring a rating of ≥85. The following offensive players are deemed stars: Sergio Agüero, Pierre-Emerick Aubameyang, Kevin De Bruyne, Edison Cavani, Bruno Fernandes, Gareth Bale, Harry Kane, Sadio Mané, Antonio Martial, Paul Pogba, Mohamed Salah, Raheem Sterling, Son Heung-min and Timo Werner.
Save	Goalkeeper saves where superlatives were used in commentary timelines (e.g. 'amazing', 'outstanding', 'exceptional' save).
Injury	A player incurs an injury leading to a significant time delay. In all instances the player was substituted.
Derby	Manchester City-Manchester United; Liverpool-Everton; Tottenham-Arsenal; Crystal Palace-Brighton and Hove Albion; Tottenham-Chelsea; Liverpool-Manchester United; West Ham-Chelsea; Tottenham-West Ham; Aston Villa-Wolves; West Brom-Aston Villa; Fulham-Chelsea.

consider foul play, which has already been investigated from the team perspective in football (Jewell, 2009) and is worth evaluating from a fan perspective.

At times, coding variables in Table 1 involves subjective judgement. For example, defining what constitutes a VAR controversy or bad mistake is open to interpretation. Given this subjectivity, significant time is dedicated to textual analysis and triangulating information from different commentary timelines and subsequent media attention. For example, we check total shots off the crossbar/post against aggregate match statistics and assessed different commentaries (e.g. BBC.com, Whoscored.com) to affirm actions such as errors or impressive saves.

For *CriticalFixture* we identify matches the media frequently promote and flag as seasonally-important on a weekly basis. By gathering the data week-to-week we gained a rich understanding of title, European qualification and relegation uncertainty. While we have no explicit criteria for this classification, seasonal important fixtures are clearly apparent as the league progressed. The method of data collection allows us a classify important league fixtures as the competition evolves; alternative teams become candidates for the Championship, European qualification and relegation as the season progresses.[6]

Other variables are less debatable such as *Derby* but still open to interpretation. The *NearMiss* variable is created through textual analysis of commentary timelines on Sky Sports. Key search terms for coding this variable include "off the bar/crossbar/post/line", "hits the bar/crossbar/post/line", "woodwork", "post", "crossbar", "bar", "goalframe" and "goal line clearance". Totals are crosschecked with aggregate match statistics available from the Premier League and websites such as the BBC and Whoscored.com.

These variables are defined in Table 2, with each data point available on request.

3.3. Model

The empirical model takes the form:

$$ln(Views)_{it} = \alpha + \beta_1 BCD_{it} + \beta_2 Shock + \beta_3 xG + \beta X_{it}^{'} + \beta Z_{it}^{'} + \beta S_{it}^{'} + \eta_{it} + .u_{it}$$

Table 3
Summary statistics - views.

Continuous Variables	Mean	SD	Min	Max
Weekly Views	596,433	363,119	39,998	2,112,590
End-of-Season Views	673,896	448,274	42,346	2,917,725
Weekly Views (Attended)	594,660	384,461	52,066	1,958,184
Weekly Views (BCD)	618,462	353,883	39,988	2,112,590
End-of Season Views (Attended)	681,864	503,779	59,868	2,917,725
End-of-Season (BCD)	691,950	413,701	42,346	2,678,395

We pursue a simple modelling strategy where the unit of observation is the natural log of YouTube views (*Views*) and denotes the outcome variable for match *i* on game-week *t* measured on a weekly basis.[7] *BCD* matches are captured by a dummy indicator that holds a value of 1 if a match is BCD and 0 otherwise.

Our *shock* measure controls for marginal goal difference in the outcome and we measure *xG* to capture match quality. The vectors $X_{it}^{'}$ and $Z_{it}^{'}$ include a series of variables to account for pre-match features and specific in-play events for match *i* on game-week *t*. $S_{it}^{'}$ is a vector representing scheduling controls for relevant aspects of match organisation such as days of the week and the level of the subscription pay-wall (basic subscription and box-office). We also control for free-to-air matches on Sky's non subscription channels.[8] We include club (match) fixed effects - η_{it} - to account for unobservable, time-invariant heterogeneity associated with both clubs involved in a fixture. u_{it} is an error term. As there is little variation in the reel lengths uploaded to YouTube (~3 min), we do not include time controls. We experimented with various interaction effects (e.g., *GOTM* and *Big6* etc.) - none of these make the final specification due to the negligible effects. Table 3 provides summary statistics for the outcome variables. Table 4 provides summary statistics for the independent variables the at a match level.

4. Results

4.1. Main models

Table 5 displays the results. The coefficients across both models are mostly consistent in statistical significance and the direction. BCD matches are associated with a significant increase in streaming views. This boost to online viewership is robust to excluding partially attended matches during the COVID period.

We observe significant coefficients ($p < 0.01$) for shock.[9] Granted, we use an alternative definition, this finding broadly complements Buraimo et al. (2020) and further emphasizes the importance of tacit characteristics of the product. In short, surprising outcomes are important for fan engagement after-the-fact.

Unsurprisingly, we report a positive effect for match xG and in-play events such as exceptional goals (*GOTM*), critical errors (*Howlers*) and penalties (*Penalty*). Foul play (*Redcard*) and star goal scorers also increase views. While the direction and statistical significance of these effects are not surprising, the result is thematic with the literature concerning team demand for aggression in sport (Jewell, 2009; Rockerbie, 2012) and the EPL literature that studies the importance of star players[10] to broadcasting demand (Buraimo & Simmons, 2015; Scelles, 2017). We find evidence to suggest that VAR incidents (*VARCon*) stimulate increased viewership. Although this relationship is only at the 10% level in Model 1, deliberations arising from the use of the technology attracted higher

[6] All weekly viewer data points are available on request.

[7] We use the log of views given the skewness in the outcome variable. This is confirmed using a Shapiro-Wilk's test for weekly views (W = 0.93, V = 33.04, Z = 8.56, Prob > Z = 0.00) and end-of-season views (W = 0.88, V = 53.75, Z = 9.75, Prob > Z = 0.00). We also checked the outcome variables for unit root problems prior to the estimations using Augmented Dickey-Fuller tests to ensure stationarity.

[8] On occasion a free sample was extended to an entire match.

[9] This finding is robust to alternative definitions of shock such as defining shock as a favourite losing only.

[10] We performed checks on the impact of stars – extending the rating bounds to include additional players and altering the mixture of 'stars' – but this did not materially impact the results.

Table 4
Summary statistics – independent variables.

Continuous Variables	Mean	SD	Min	Max
Shock (Attended)	9.077	4.159	0.25	17.21
Shock (BCD)	9.107	4.577	0.33	19.16
xG (Attended)	2.915	0.999	0.28	6.97
xG (BCD)	2.704	1.021	0.78	6.41
NearMiss (Attended)	0.537	0.712	0	3
NearMiss (BCD)	0.750	0.909	0	5
Game-week (Attended)	16.257	9.688	1	38
Game-week (BCD)	21.769	11.121	1	38

Indicator Variables	No. of Obs. (Proportion) Attended	No. of Obs. (Proportion) BCD
Big6	25 (7.86%)	35 (8.08%)
Derby	14 (4.04%)	25 (5.77%)
NewManager	10 (3.14%)	5 (1.15%)
CriticalFixture	11 (3.46%)	89 (20.55%)
StarScore	81 (25.47%)	111 (25.64%)
Hat-Trick	7 (2.20%)	13 (3.00%)
VarCon	67 (21.07%)	88 (20.32%)
RedCard	38 (11.95%)	47 (10.85%)
Injury	25 (7.86%)	22 (5.08%)
Howler	31 (9.75%)	39 (9.01%)
Save	31 (9.75%)	13 (3.00%)
GOTM	54 (16.98%)	75 (17.32%)
Penalty	69 (21.70%)	106 (24.48%)
Holiday	30 (9.43%)	27 (6.24%)
Friday	9 (2.83%)	18 (4.16%)
Saturday	169 (53.14%)	130 (30.32%)
Sunday	79 (24.84%)	131 (30.25%)
Monday	10 (3.14%)	45 (10.39%)
Midweek	51 (16.04%)	109 (25.17%)
SubTV	154 (48.43%)	390 (90.07%)
BoxOffice	0 (0%)	14 (3.32%)
FTATV.	0 (0%)	34 (7.85%)

levels of stream.

Several pre-defined match features do not impact weekly views. Rivalries (*Derbies*) are a salient example. Caution is needed when interpreting the impact of pre-match factors however as fans may be motivated by these to watch matches live. This logic could be a possible explanation for the negative *Big6* effect we observe. Typically, these are the most advertised and popular matches. As this is somewhat a counter-intuitive finding, we explore whether this is a function of collinearity problems with broadcasting variables. The negative *Big6* persists without the broadcasting variables. As we cannot account for live viewership, the inferences one can make from these characteristics - which represents known information to fans - is limited.

Finally, we find a range of scheduling effects. Relative to midweek, Friday and Monday matches are viewed more, while Saturday matches are viewed less. We conjecture that Friday content could be favoured as this is the first match in a game-week and fans are eager for fresh fixtures. Fans may also have social or travel commitments on a Friday night that are not present during daytime kick-offs on Saturday or Sunday and are therefore more likely to miss a game. The positive Monday effect may be explained by commitments that are not present on weekends. The negative Saturday effect is resilient to normalising the dependent variable. This effect may be owing to competing broadcasts that show the same content (e.g., Match of the Day) which are scheduled on a Saturday evening or consumers having greater time to watch sport live on this day and thereafter not holding a preference to review watched content. We identify an anticipated non-linear impact of game-week on streaming. Finally, and expectedly, *BoxOffice* matches – placed behind a double pay wall – where an additional type of once-off subscription was required attracted additional streams.

4.2. Checks

We estimate various additional regressions to consider the robustness of the results for our preferred weekly model. For instance, we implemented auxiliary methods that cope with the skewed outcome variable. We estimate the parameters of a standard (Gaussian) generalised linear regression model (GLM) using maximum likelihood optimization. This approach, along with both median regression and robust regression, produce similar effects to those observed in Table 5 (see appendix table 6).

Table 5
Weekly & end-of-season views.

Coefficient	MODEL 1 – Weekly Views		MODEL 2 – End-of-Season Views	
	ln(Views)	RSE	ln(Views)	RSE
BCD	0.129***	(0.038)	0.084**	(0.038)
Shock	−0.041***	(0.003)	−0.044***	(0.003)
xG	0.089***	(0.018)	0.090***	(0.017)
Big6	−0.389***	(0.068)	−0.303***	(0.066)
Derby	0.023	(0.067)	0.046	(0.067)
NewMan	−0.096	(0.081)	−0.121	(0.088)
CriticalFixture	0.002	(0.054)	0.003	(0.053)
StarScore	0.163***	(0.046)	0.201***	(0.041)
NearMiss	−0.006	(0.018)	−0.002	(0.017)
Hat-trick	0.093	(0.080)	0.104	(0.084)
VARCon	0.061*	(0.037)	0.089**	(0.034)
RedCard	0.134***	(0.051)	0.131**	(0.051)
Injury	0.026	(0.063)	0.017	(0.061)
Howler	0.112**	(0.049)	0.127***	(0.048)
Save	−0.023	(0.062)	−0.022	(0.064)
GOTM	0.242***	(0.037)	0.251***	(0.035)
Penalty	0.074**	(0.034)	0.058*	(0.034)
Game-week	0.022***	(0.006)	0.022***	(0.006)
Game-week^2	−0.0005***	(0.001)	−0.0006***	(0.001)
Holiday	−0.060	(0.053)	−0.062	(0.051)
Friday	0.247***	(0.069)	0.152**	(0.068)
Saturday	−0.084*	(0.045)	−0.188***	(0.041)
Sunday	0.052	(0.049)	−0.047	(0.045)
Monday	0.164**	(0.065)	0.084	(0.063)
SubTV	0.010	(0.048)	0.015	(0.046)
BoxOffice	0.189**	(0.092)	0.178*	(0.092)
FTATV	−0.056	(0.084)	0.010	(0.079)
Constant	12.228***	(0.186)	10.785***	(0.172)
R^2	0.661		0.694	
N	751		751	
Prob > F	0.000		0.000	
F-Stat	33.19		37.96	
VIF	2.87		2.87	

*p < 0.10; **p < 0.05; ***p < 0.01. Robust standard errors in parentheses. Club fixed effects included. Base Indicator: Midweek.

We redefined our indicator variable for BCD so that limited attendances are omitted. This is a worthwhile investigation as stadia were not completely closed for all BCD matches. The number of fans permitted to enter was however relatively low; the reported attendances for these limited attendances only ranged from 2,000 to 10,000 fans. As expected, tweaking our definition of a BCD match does not alter the statistical significance of the results (see appendix table 7).

One may be concerned that highlights are hosted only from August 2019 onwards and thus the BCD dummy may simply capture a trend. We do not control for any marketing-related activities, promotion or diffusion effects. To address this, we analysed a restricted sample from game-week 6 onwards. This begins on 20th September 2019 allowing for one month of marketing or general diffusion effects to take hold from the beginning of the season. The BCD variable remains statistically significant at the 5% level (0.084, RSE = 0.039, P = 0.034).

5. Discussion & conclusion

We explore fan preferences to stream content in light of BCD matches caused by COVID-19 and depart from the traditional ex ante approach to modelling football viewership. Furthermore, we offer the first study to consider ex post demand via a video sharing application. As our results reveal specific characteristics of football that fans are interested in, we effectively look under the hood of broadcasting demand. This is timely as football highlights and other low-end versions of content, are now commonly offered for free online, via legitimate third-party platforms.

Our insights carry various practical recommendations. The purpose of giving away free content online - as is the case in our context - is to transition a proportion of free/low-end users to subscriptions. Our results thus aid broadcasters in maximising the benefits of the low-end product as we identify traits that attract higher views. At present, content via our data source is presented using a basic chronological strategy. This appears to be the case across a variety of YouTube sports channels for major broadcasters. Our results can inform this strategy. For example, matches with shock outcomes ought to be positioned prominently on a menu or streaming feed. For shorter digital promotion through social media, firms should focus on advertising with the dichotomous in-play effects identified. For example, hosting a short clip of a bad mistake or red card would be more likely to garner attention than a clip of an impressive save or near miss. This implication is also relevant to public relations and communications teams within football clubs aiming to increase fan engagement.

Several limitations of our study are noteworthy. First, a possible confound in our analysis is the presence of (social)media and its influence on viewership. This could only be loosely mitigated (i.e., by assessing tweet volume or media mentions), however this would raise obvious collinearity concerns when specifying the underlying in-play measures in a regression model. We are also concerned that using social media data that is not clean could bias the results. For this reason, we did not assess the role of electronic word-of-mouth, but note this as a limitation, as it is possible that media coverage could affect views.

Second, we are assuming that all other material created and distributed by the Sky Sports channel is random – it may not be. The fact that match highlights are posted chronologically however, without any strategic consideration, does lend itself to the view that no detailed and explicit content creation strategy exists.

Third, while there are many advantages to our dataset, as it was collected over a relatively long time, it is possible that exogenous factors could shift views over the two seasons and are not captured empirically. For instance, the emergence of substitute products, competing sports channels or general competition effects in the broadcasting space could potentially impact viewership trends. Moreover, YouTube algorithms could adapt to consumer preferences and nudge viewers toward selective content. While the YouTube channel we consider is a standardised format, the videos from the channel could spill over to other user feeds. Despite these limitations the findings prompt questions for future research. In particular, future research could consider fan BCD viewing preferences through exploring other football competitions and different sports to question the generality of the findings.

Appendix Table 6
Weekly Views - Alternative Estimators

Coefficient	GLM		Huber/Robust Regression		Quantile (Median) Regression	
	Views	RSE	Views	SE	Views	RSE
BCD	54,274**	(21,352)	61,315***	(19,208)	59,645***	(17,802)
Shock	−24,049***	(1,999)	−22,172***	(1,877)	−21,484***	(1,559)
xG	36,083***	(9,285)	36,262***	(8,585)	41,575***	(6,747)
Big6	−112,388**	(45,780)	−70,231*	(36,544)	−57,097	(43,725)
Derby	33,310	(50,781)	36,292	(35,353)	13,495	(47,247)
NewMan	−108,888***	(39,182)	−94,735*	(53,186)	−150,600**	(58,708)
CriticalFixture	−5,288	(29,137)	16,260	(25,864)	14,786	(25,278)
StarScore	139,079***	(27,549)	119,982***	(22,543)	135,197***	(24,849)
NearMiss	1,211	(9,205)	−4,750	(9,065)	−16,246**	(7,415)
Hat-trick	137,587**	(64,203)	165,028***	(47,711)	155,395***	(45,722)
VARCon	28,595	(21,491)	41,475**	(18,919)	37,558*	(19,368)
RedCard	90,444***	(27,196)	59,964**	(23,384)	64,137***	(24,474)
Injury	29,769	(42,827)	−6,597	(31,039)	21,160	(40,169)
Howler	94,624***	(33,772)	77,169***	(25,766)	106,287***	(36,415)
Save	−30,193	(30,366)	−15,986	(31,405)	−37,280	(28,028)
GOTM	136,480***	(22,831)	119,609***	(19,546)	121,967***	(19,927)
Penalty	44,250**	(20,196)	45,757**	(18,701)	27,580**	(13,934)
Game-week	10,952***	(3,211)	10,608***	(2,958)	10,396***	(2,979)
Game-week^2	−281***	(82)	−283***	(76)	−269***	(72)
Holiday	−45,203*	(26,786)	−22,322	(28,960)	−5,546	(29,477)
Friday	95,736**	(41,425)	87,821**	(42,976)	81,013	(53,204)
Saturday	−56,215**	(22,806)	−40,807*	(21,108)	−46,602***	(17,692)
Sunday	23,474	(26,633)	21,054	(22,091)	6,095	(21,601)
Monday	59,903*	(30,769)	64,502*	(32,880)	64,151***	(24,261)
SubTV	−28,300	(22,455)	−9,349	(21,013)	9,941	(17,472)
BoxOffice	131,466**	(63,096)	149,452**	(58,507)	103,247	(100,868)
FTATV	−49,676	(36,409)	−42,237	(39,020)	−44,374	(27,710)
Constant	270,019***	(83,078)	206,730***	(78,035)	167,883**	(83,249)
N	751		751		751	

Appendix Table 7
Weekly & end-of-season views – reduced attendance check.

Coefficient	MODEL 1 – Weekly Views		MODEL 2 – End-of-Season Views	
	ln(Views)	RSE	ln(Views)	RSE
BCD	0.097**	(0.038)	0.119***	(0.036)
Shock	−0.042***	(0.004)	−0.045***	(0.003)
xG	0.081***	(0.018)	0.084***	(0.018)
Big6	−0.388***	(0.067)	−0.304***	(0.065)
Derby	0.019	(0.066)	0.042	(0.065)
NewMan	−0.114	(0.083)	−0.134	(0.089)
CriticalFixture	0.024	(0.054)	0.019	(0.052)
StarScore	0.172***	(0.047)	0.207***	(0.041)
NearMiss	−0.001	(0.018)	0.001	(0.017)
Hat-trick	0.110	(0.083)	0.115	(0.084)
VARCon	0.056	(0.037)	0.084**	(0.035)
RedCard	0.133**	(0.052)	0.128**	(0.051)
Injury	0.006	(0.064)	−0.001	(0.061)
Howler	0.117**	(0.049)	0.121**	(0.049)
Save	−0.057	(0.059)	−0.050	(0.061)
GOTM	0.241***	(0.038)	0.250***	(0.036)
Penalty	0.084**	(0.034)	0.065*	(0.034)
Game-week	0.029***	(0.007)	0.030***	(0.007)
Game-week^2	−0.001***	(0.000)	−0.001***	(0.000)
Holiday	−0.055	(0.053)	−0.057	(0.052)
Friday	0.231***	(0.071)	0.142**	(0.067)
Saturday	−0.101**	(0.045)	−0.202***	(0.041)
Sunday	0.043	(0.048)	−0.055	(0.044)
Monday	0.159**	(0.066)	0.078	(0.064)
SubTV	0.095**	(0.045)	0.086**	(0.043)
BoxOffice	0.253***	(0.093)	0.211**	(0.092)
FTATV	−0.025	(0.087)	0.017	(0.080)
Constant	12.152***	(0.191)	12.323***	(0.182)
R^2	0.659		0.697	
N	751		751	
Prob > F	0.000		0.000	
F-Stat	32.55		39.03	
VIF	2.97		2.97	

*p < 0.10; **p < 0.05; ***p < 0.01. Robust standard errors in parentheses. Club fixed effects included. Base Indicator: Midweek.

Declaration of competing interest

This research did not receive any grants from funding agencies in the public, commercial or not-for-profit sectors. We also have no conflict of interest or competing interests to declare.

Acknowledgements

The authors wish to thank attendees at the National University of Ireland, Maynooth Department of Economics Seminar Series in September 2021, Gijon XV Conference on Sports Economics: Roger Noll and Peter Sloane Golden Anniversary December 2021 and two anonymous reviews for helpful comments and suggestions.

References

Bbc. (2018). Match of the day: BBC extends premier league highlights deal (2018, jan 30) BBC. Available at: https://www.bbc.com/sport/football/42873526.

Borland, J., & MacDonald, R. (2003). Demand for sport. *Oxford Review of Economic Policy, 19*(4), 478–502.

Bryson, A., Dolton, P., Reade, J. J., Schreyer, D., & Singleton, C. (2021). Causal effects of an absent crowd on performances and refereeing decisions during Covid-19. *Economics Letters, 198*, Article 109664.

Buraimo, B., Forrest, D., McHale, I. G., & Tena, J. D. (2020). Unscripted drama: Soccer audience response to suspense, surprise, and shock. *Economic Inquiry, 58*(2), 881–896.

Buraimo, B., Forrest, D., McHale, I. G., & Tena, J. D. (2021). Armchair fans: Modelling audience size for televised football matches. *European Journal of Operational Research*.

Buraimo, B., & Simmons, R. (2008). Do sports fans really value uncertainty of outcome? Evidence from the English premier league. *Int. J. Sport Financ., 3*(3), 146–155.

Buraimo, B., & Simmons, R. (2015). Uncertainty of outcome or star quality? Television audience demand for English premier league football. *International Journal of the Economics of Business, 22*(3), 449–469.

Butler, R., & Massey, P. (2019). Has competition in the market for subscription sports broadcasting benefited consumers? The case of the English premier league. *Journal of Sports Economics, 20*(4), 603–624.

Caselli, M., Falco, P., & Mattera, G. (2021). *When the stadium goes silent: How crowds affect the performance of discriminated groups*.

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

Cross, J., & Uhrig, R. (2023). Do fans impact sports outcomes? A COVID-19 natural experiment. *Journal of Sports Economics, 24*(1), 3–27.

De Angelis, L., & Reade, J. J. (2022). Home advantage and mispricing in indoor sports' ghost games: The case of European basketball. *Annals of Operations Research*, 1–28.

Farnell, A. (2023). False start? An analysis of NFL penalties with and without crowds. *Journal of Sports Economics*.

Ferraresi, M., & Gucciardi, G. (2021). Who chokes on a penalty kick? Social environment and individual performance during covid-19 times. *Economics Letters, 203*.

Feuillet, A., Scelles, N., & Durand, C. (2019). A winner's curse in the bidding process for broadcasting rights in football? The cases of the French and UK markets. *Sport in Society: Cultures, Commerce, Media, Politics, 22*(7), 1198–1224.

Fischer, K., & Haucap, J. (2021). Does crowd support drive the home advantage in professional football? Evidence from German ghost games during the COVID-19 pandemic. *Journal of Sports Economics, 22*(8), 982–1008.

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

Gong, H. (2022). The effect of the crowd on home bias: Evidence from NBA games during the COVID-19 pandemic. *Journal of Sports Economics*.

Jewell, R. T. (2009). Estimating demand for aggressive play: The case of English premier league football. *Int. J. Sport Financ., 4*(3), 192–210.

Jewell, R. T. (2017). The effect of marquee players on sports demand: The case of US Major League Soccer. *Journal of Sports Economics, 18*(3), 239–252.

Jones, A. (2019). Turning 'nuggets into gold', rows with everyone and the 3pm TV blackout – how 'Lord of Burnley' shaped the game. (2019, Oct 25) the Athletic. Available at: https://theathletic.com/1317306/2019/10/25/profiting-from-youngsters-better-facilities-than-chelsea-the-3pm-tv-blackout-banning-reporters-how-lord-of-burnley-shaped-the-game/.

Kharrat, T., McHale, I. G., & Peña, J. L. (2020). Plus–minus player ratings for soccer. *European Journal of Operational Research, 283*(2), 726–736.

Leitner, M. C., Daumann, F., Follert, F., & Richlan, F. (2022). The cauldron has cooled down: A systematic literature review on home advantage in football during the COVID-19 pandemic from a socio-economic and psychological perspective. *Manag. Rev. Quart.*, 1–29.

Losak, J. M., & Sabel, J. (2021). Baseball home field advantage without fans in the stands. *Int. J. Sport Financ., 16*(3).

Reade, J. J., Schreyer, D., & Singleton, C. (2021). Stadium attendance demand during the COVID-19 crisis: Early empirical evidence from Belarus. *Applied Economics Letters, 28*(18), 1542–1547.

Reade, J. J., Schreyer, D., & Singleton, C. (2022). Eliminating supportive crowds reduces referee bias. *Economic Inquiry, 60*(3), 1416–1436.

Rockerbie, D. W. (2012). The demand for violence in hockey. In L. H. Kahane, & S. Shmanske (Eds.), *The oxford handbook of sports economics: Volume 1: The economics of sports* (pp. 159–175). New York: Oxford University Press.

Scelles, N. (2017). Star quality and competitive balance? Television audience demand for English premier league football reconsidered. *Applied Economics Letters, 24*(19), 1399–1402.

Scelles, N., Dermit-Richard, N., & Haynes, R. (2020). What drives sports TV rights? A comparative analysis of their evolution in English and French men's football first divisions, 1980-2020. *Soccer and Society, 21*(5), 491–509.

Smith, M. (2004). *Match of the day: 40th anniversary*. London): BBC Books.

Szymanski, S. (2003). The assessment: The economics of sport. *Oxford Review of Economic Policy, 19*(4), 467–477.

Winkelmann, D., Deutscher, C., & Ötting, M. (2021). Bookmakers' mispricing of the disappeared home advantage in the German Bundesliga after the COVID-19 break. *Applied Economics, 53*(26), 3054–3064.

Wolaver, A. M., & Magee, C. (2022). Ghost games: Crowds, referee bias, and home advantage in European football leagues. *Journal of Sport Behavior, 45*(3).

Yang, Y., & Shi, M. (2011). Rise and fall of stars: Investigating the evolution of star status in professional team sports. *International Journal of Research in Marketing, 28*(4), 352–366.