Does certainty on the winner diminish the interest in sport competitions? The case of Formula One

Article

It is advisable to refer to the publisher's version if you intend to cite from the work. See Guidance on citing.

To link to this article DOI: http://dx.doi.org/10.1007/s00181-021-02147-8

Publisher: Springer

www.reading.ac.uk/centaur

CentAUR

Central Archive at the University of Reading

Reading's research outputs online

Does certainty on the winner diminish the interest in sport competitions? The case of formula one

Pedro Garcia-del-Barrio[1] · J. James Reade[2]

Abstract

The literature acknowledges "Uncertainty of Outcome" (UO) as a major factor to explain the degree of interest that sporting competitions draw from fans and the general public. Uncertainty about the championship winner is crucial insofar as financial success depends on the capacity to attract potential consumers of spectacle. This paper focusses on one aspect of UO and examines to what extent reduction in the interest of followers is due to the removal of uncertainty about the world drivers' champion in Formula One. To study how certainty on the winner undermines the degree of attention generated by the Formula One world drivers' championship, we rely on two alternative indexes—similar although not identical—reported by *Google Trends*. Both of these appraisals are computed from data on users' search intensity in Google, where weekly records are normalised on the relative amount of searches per calendar year. Thus, as dependent variables for the empirical analysis we use two measures: *Google Trends News* (GTN), to capture the intensity with which individuals search news articles associated; and *Google Trends Web* (GTW), to get a wider overview based on all kind of Internet contents. The former empirical analysis is carried out on 10 years of available data, while the latter approach estimates the models for a larger period of 14 years. Our empirical strategy includes additionally adopting indicator saturation techniques to address this issue while controlling for outliers.

✉ Pedro Garcia-del-Barrio
pgbarrio@unav.es

J. James Reade
j.j.reade@reading.ac.uk

[1] Edificio Amigos, Universidad de Navarra, Campus Universitario, 31009 Pamplona, Navarra, Spain

[2] University of Reading, Building: Edith Morley 321a, Reading, Berkshire RG6 6AH, UK

🖉 Springer

1 Introduction

Entertainment industries are developed upon the surprise and suspense attached to unknown outcomes. The value created by this type of businesses largely depends on the capacity to attract audiences, which—in turn—depends on the degree of uncertainty and mystery associated with the final result. The sports economics literature has long been dominated by the *Uncertainty of Outcome* (UO) hypothesis, which advocates that the interest in a sporting event depends on the outcome being sufficiently uncertain. Academic debate has centred on how much uncertainty is necessary, but by and large has failed to draw clear conclusions regarding the existence of support for the hypothesis. This paper must actually be placed in the context of the championship uncertainty (as different from match or race uncertainty). The issue is crucial since it relates to the sport competitions' capacity to attract attention and to generate economic returns.

Most of the literature on the Uncertainty of Output Hypothesis (UOH) focusses on the uncertainty attached to a single match or race (Fort and Quirk 1995). Previously, the seminal paper by Whitney (1988) stresses the effect of the absence of uncertainty about the season champion. The former paper (page 1267 to 1268) claims that their *"win percent"* model overperforms the latter model. However, there are few papers to help elucidating which of the two positions has stronger empirical support (Humphreys and Zhou 2015).

By distinguishing the degree of race (or match) uncertainty from championship uncertainty, the scope of this paper embraces both sources of uncertainty, and empirically examines the existing tension between them. Based on records gathered from Google Trends, our empirical analysis is potentially valid to evaluate the extent to which each of these two approaches finds greater empirical support in the context of the Formula one. Our results actually acknowledge that both approaches (Whitney/Fort and Quirk) find certain degree of support in the context of Formula One. Hence, this paper aims to fulfil a gap in the literature, by empirically examining the distinction between match and championship uncertainty.

Some of the more interesting insights have been drawn from considering uncertainty from the perspective of fans (rather than from a statistical or econometric perspective). Useful insights flow from the realisation that uncertainty exists at different levels in any sporting contest; there is the individual event level (a match, or race), the seasonal level (the pennant race), and the multi-seasonal level (long-term domination). Given this multi-dimensional nature to uncertainty of outcome, it may thus be possible to consider the impact of removing one aspect of uncertainty on interest levels.

In line with this angle on the question, we examine what reduction in interest is associated with the removal of uncertainty on the season winner. Most sporting events have what Neale (1964) referred to as a pennant race—the competition for

the ultimate prize. Outside of the major American sports, that ultimate prize can be determined before all matches in the competition have been completed. Progression from round-robin tournaments can be known before the final round matches take place. Almost all sporting contests involve such a pennant race, along with individual events too. Besides, the removal of pennant race uncertainty allows the identification of the interest levels in the uncertainty of individual events. The issue is relevant concerning the signing of sponsorship contracts and potential access to new fans' crowds, as the degree of interest attached to one sport competition (relative to other alternative competitions) may be eroded when the winner is know.

There are diverse communication channels to follow sporting competitions, from the more traditional media (TV, radio, newspapers, etc.) to more modern devices (Internet, mobile apps, social networks, etc.). The internet undoubtedly dominates; in 2005, the International Telecommunications Union estimates that 51% of the developed world had regular internet access, and in 2019 that figure was 86.6%.[1] The standard mechanism for finding information on the internet is to search using a search engine, and of search engines, Google has overwhelmingly dominated for well over a decade; in January 2009 GlobalStat estimated 90.25% of web search traffic using Google, and in February 2020 that number had edged up slightly to 92%.[2] Thus, one way to measure the degree of fan interest is to consider how often people search for information about that sport on Google. It is to be expected that greater uncertainty yields greater interest, which can be corroborated by comparing the number of searches made by Google's users.

In this paper, we hypothesise that interest falls once the world drivers' champion is known. Moreover, given that our empirical study is realised from a global perspective, we venture that studying the impact of (un)certainty on the championship's winner may help predicting the future economic prospects of sport competitions.

After introducing in Sect. 1 the topic, Sect. 2 reviews the most relevant previous literature. In Sect. 3, we describe the data set and sources and set out the adopted modelling methodology. Then, Sect. 4 presents the main results from the econometric estimations; and, finally, Sect. 5 concludes.

2 Related literature

The discussion on unpredictable outcomes, since the early stages of sports economics as a discipline, has been closely linked to the issue of competitive balance (or unbalance). Rottenberg (1956) is well known for arguing that more intense competition is positively related to the level of interest of fans—as captured by attendances. Since the publication of his seminal paper, Rottenberg's conjecture is known as the *Uncertainty of Outcome Hypothesis* (UOH). A few years later, Neale (1964) referred to the UOH as a factor to explain the fans' degree of interest, even if this feature is only mentioned in a footnote, where we read: *"the appeal of the seat depends mostly on the uncertainty of outcome and on the weather"*. Fort and Quirk (1995) conducted a survey taking for granted the recognition of the difference between *competitive balance* (CB) and the

---

[1] See https://en.wikipedia.org/wiki/List_of_countries_by_number_of_Internet_users.

[2] See https://gs.statcounter.com/search-engine-market-share.

league-standing effect, the notion that better quality teams attract more fan interest as well. Moreover, the relationship between CB and UO has continued to stimulate new studies (Owen 2014 and Késenne (2014), for instance).

Given the difficulty to measure CB in professional team-sport leagues, research efforts tried to elucidate how this task could be performed. Concerning the studies on the effect that CB exerts on the degree of interest in sport competitions, there are contrasting empirical results. While some papers find a significant positive impact (mainly on stadium attendances), others reach the opposite conclusion (Peel and Thomas 1992; Czarnitzki and Stadtmann 2002; García and Rodríguez 2002; Borland and MacDonald 2003; Lee and Fort 2008; Coates and Humphreys 2010, 2012; Manasis et al. 2013, among others).[3]

Zimbalist (2002) observed that there is a general acceptance on the fact that leagues must enjoy certain CB, but also that economists do not know how much of it should be preferred. Owen (2014, page 41) summarises some major aspects concerning the Rottenberg's UOH: *"According to the UOH, a higher degree of predictability (that is, reduced uncertainty) erodes spectators' interest. This decreases consumer demand for the league's output, reflected in lower match attendances, lower television audience ratings and reduced opportunities for sponsorship and merchandising. Amongst economists, it is not a widely held view that the 'ideal' degree of CB involves perfect balance, with all teams of equal strength".*

Based on previous studies (Cairns et al. 1986; Szymanski 2003; and Kringstad and Gerrard 2007),[4] Késenne (2014) differentiates three levels of UO in sports: (i) match uncertainty; (ii) seasonal or "within-season" uncertainty (a dynamic measure which takes into account more than one season); (iii) and championship or "between-season" uncertainty. Cairns et al. (1986) actually argued that uncertainty of outcome matters at three levels: the match or individual event level, the seasonal or championship level (who wins the pennant race, as Neale (1964) described it), and the multi-seasonal level (dominance over the years). Indeed, each one of these matters, and in some cases, there are multiple dimensions within-season. Our analysis focusses on the degree of interest generated by sport events, regardless of whether obtaining information provides entertainment or not (Ely et al. 2015).[5]

---

[3] The issue is a matter of debate in the context of European football: Késenne (2000) or Szymanski (2001); and in Formula One: Mastromarco and Runkel (2009) or Judde et al. (2013). Other papers adopt an approach that examines TV audiences (Perez et al. 2017).

[4] Then, Jennett (1984) introduced a within-season measure of uncertainly taking into account if the teams were still candidates for winning the championship and also the number of games left for the competition to be completed. Humphreys (2002) use annual figures on attendances to baseball stadiums to examine alternative measures of CB, suggesting an innovative single measure of UO that combines seasonal and championship uncertainty: the *"competitive balance ratio"*. More recently, Pawlowski and Anders (2012) adopts a new approach to overcome the usual shortcomings associated with the appraisal of CB by means of subjective measurements of fans' perceptions. Manasis et al. (2013) create one such specialised index for football leagues accounting for promotion and relegation.

[5] Ely et al. (2015) elaborate on a further refinement of the uncertainty concept, by distinguishing between the entertainment utility derived from suspense and the one stemming from surprise: *"A period has more suspense if the variance of the next period's beliefs is greater. A period has more surprise if the current belief is further from the last period's belief", according to their definitions"* (page 216). Buraimo et al. (2020) use an extensive data set on television audiences to address the issue of outcome uncertainty, concluding that suspense and surprise are significant drivers of fans' interest on Premier League matches.

3 Data and methodology

To measure the degree of interest generated by sporting competitions, we examine how often fans search on the Internet for information about the Formula One championship, an sport discipline that awakes the attention of large crowds worldwide (Garcia-del-Barrio 2018). This approach seems appropriate in the Internet age, when the mechanism for finding more information out about something is to conduct a web search. If a sporting event is interesting, either if one is a current fan or not, it is plausible that they may search for information about the event using a search engine. Overwhelmingly over this time period, Google has been the dominant search engine globally. Consumers of sport spectacles will actually make Google searches to either find websites of interest, or news articles on the issue that has piqued their interest.[6]

We collected weekly figures of searches in the *Google Trends* tool as a proxy variable to measure the degree of attention paid by the consumers of sports spectacle. The *Google Trends* algorithms identify and label the searches linked to Formula One. For the sake of robustness in the results, we actually use data on both the intensity with which users searched on Google for general web content (*Google Trends Web*, GTW) and for news articles (*Google Trends News*,GTN). The latter approach permits appraising the awareness of people concerning news articles related to the sport competitions, whereas the former captures all kinds of Internet searches, thereby providing a more global view on the ability that sport competitions have to draw the attention of the fans and the general public. The data are downloaded from the *Google Trends* site.[7] The web searches (GTW) are plotted in Fig. 1, while the news searches (GTN) are plotted in Fig. 2, where seasonal patterns in the searches are clearly observed. At the start and end of the year, when the Formula One season is not running (the week of the year is on the horizontal axis), search volumes are low. The same is true for the mid-season break, clearly visible just after week 30. In each of the two alternative approaches, GTW and GTN, we examine the comparative intensity in searching information on the Formula One world drivers'

---

[6] As a measure, it circumvents the problem of endogeneity that would be a concern when looking at, say, newspaper column inches, since this is a decision made by newspaper editors as a calculation based on anticipated interest in such content; editors make the decision because they anticipate the demand, rather than because the demand is there. There are several papers using the intensity of searches in *Google Trends* to approximate the degree of interest manifested by fans and the general public (Garcia-del-Barrio et al. 2020), Aguiar-Noury and Garcia-del-Barrio (2019) and (2021). Previous studies (Vosen and Schmidt, 2011) show that *Google Trends* is reliable and helps to forecast the consumers" tendencies; actually, their *"results show that in almost all conducted in-sample and out-of-sample forecasting experiments the Google indicator outperforms the survey-based indicators. This suggests that incorporating information from Google Trends may offer significant benefits to forecasters of private consumption"*. Choi and Varian (2012) claim that *Google Trends* is *"a real-time daily and weekly index of the volume of queries that users enter into Google, [which] may be helpful for short-term economic prediction. (...) Google Trends may help in predicting the present"*. Then, Bulut (2018) concludes that using figures based on from *Google Trends* overperforms other approaches in forecasting changes in the nominal exchange rates of OECD countries.

[7] See https://trends.google.com/.
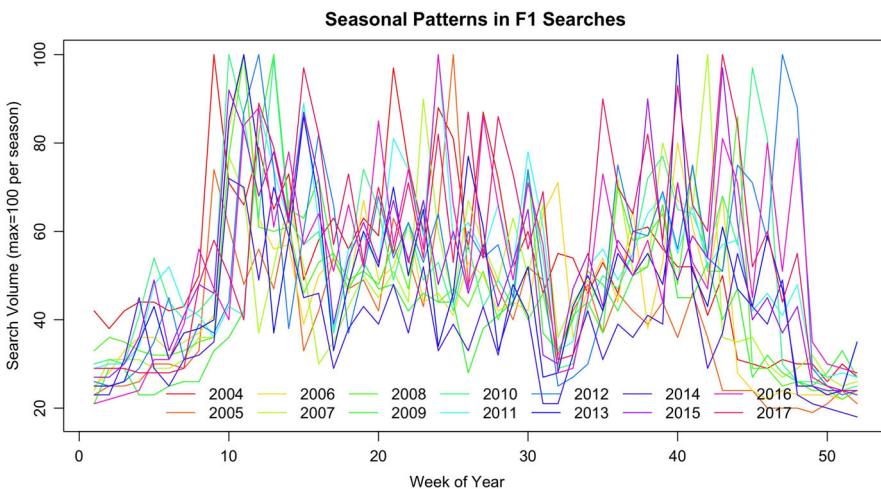
Fig. 1 Google Trends Web (GTW) searches for Formula One by week of the year, 2004–2017
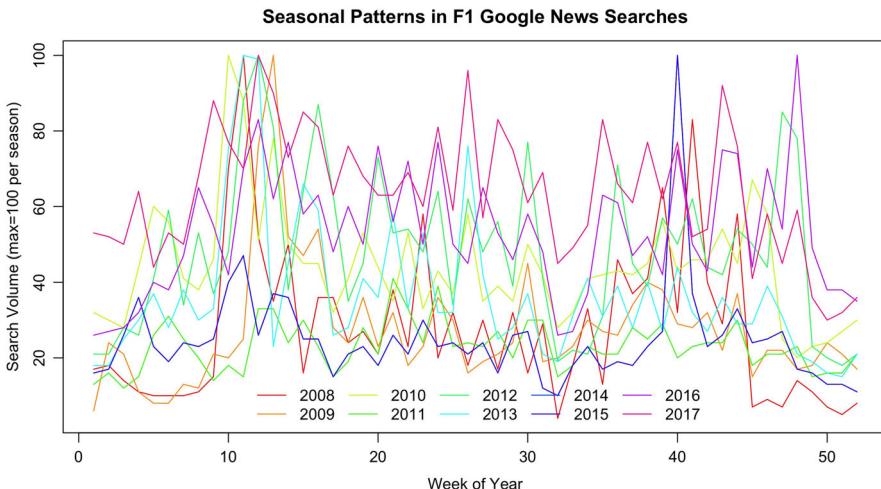
Fig. 2 Google Trends News (GTN) searches for Formula One by week of the year, 2008–2017

championship, to capture oscillations on the attention granted by the fans over the season.[8]

In principle, there is no reason to expect that the fans of sport competitions carry out a more (or less) intensive use of Internet search engines than other people do. Hence, we are confident that our analysis will deliver representative outcomes. Besides, in processing the information, we merely rely on counting the relative amount of mentions or news articles, while neglecting to account for the actual content of these Internet contents.

---

[8] Data on the race dates were collected from the Formula One official site: www.formula1.com/ and from: www.4mula1.ro.

To estimate the different models, we use the two aforementioned alternative variables. First, the dependent variable GTW is used to perform model estimations with data for a period of 14 years (from 2004 to 2017). Then, we replicate the analysis for GTN, covering just 10 seasons (from 2008 to 2017), due to shorter data availability. Hence, the length of the data sets differs for the two groups of model estimations: the former consists of 728 observations, while the latter uses 520 (the corresponding 52 annual weeks multiplied by the number of years).

Notice that Google scales its search volumes such that 100 is the maximum value for a given time period. As we collected the figures separately for each season, the series achieves 100 once per season, where seasons run for a calendar year. Thus, the records of both alternative dependent variables (GTW and GTN) are expressed relative to a maximum reference value of 100. Precisely, the use of normalised data is apropos for capturing the relative intensity of the searches made by Google users (Cf.: Choi and Varian 2012). Nonetheless, despite this implicit controlling for secular trends, it may be useful to add annual fixed effects.

In addition to these variables, we employ indicator saturation (IS) as a method of outlier detection (Cf. Hendry et al. 2013). The IS technique is a flexible method that can be applied to detecting outliers (via impulse indicator variables), structural changes (by step indicator variables), and split trends (via trend indicator variables). It works by saturating the regression with indicator variables of a particular type, and selecting over them in batches. Castle et al. (2008) finds that IS dominates other break detection procedures such as that proposed in Bai and Perron (2003), and also other conventional outlier detection measures like identifying large residuals.

Indicator saturation using steps (SIS) could have been implemented on this sample, but the use of annual fixed effect dummies interacted with the step indicators in such a way that the two counteracted each other. Furthermore, with the implicit annual detrending via the collection method from Google, changes in the level of search activity do appear to have been taken into account. What matters, particularly given this method of detrending employed by Google, is detecting for outliers—weeks where search was abnormally high for no obvious, systematic, reasons. If a huge outlier exists, it can weigh down the remaining observations and create a false sense that that season had lower search volumes.

Particularly, we aim at accounting for all persistent changes in the time series that are not attributable to knowledge of the winner. Once the final outcome (namely, who is the Formula One world drivers' champion) has become certain, we measure by how much does fan interest fall down.

Our regression-based approach involves estimating several linear models of the following form:

$$y_t = \alpha_0 + \delta \cdot D_t + \rho \cdot X_t + \alpha \cdot WinnerKnown_t + \varepsilon_t, \qquad \varepsilon_t \sim N(0, \sigma^2) \quad (1)$$

Here, $WinnerKnown_t$ is a variable taking the value 1 for the weeks when uncertainty related to the pennant race has concluded, as it is already known who has won the Formula One drivers championship. As such, it gives some estimate of the volume of interest purely associated with the pennant race.

Table 1 Descriptive statistics of the main variables

Variable	Obs.	Mean	Std.Dev.	Min	Max
Google Trends News	520	37.4846	21.0279	4	100
Google Trends Web	728	48.9560	18.7904	18	100
Racing Season Ongoing	728	0.6758	0.4683	0	1
Winner Known	728	0.0329	0.1786	0	1

In order to try and isolate pennant uncertainty, we incorporate X_t and D_t, two matrices of variables. The first, X_t, collects other explanatory variables, and the second, D_t, contains indicator variables of varying types.

Throughout the season, a number of particular weeks are of great interest, and hence are included in X_t:

1. The week before the season begins, when a huge amount of information will be put out in news media to appeal to Formula One racing fans anticipating the new season.
2. The week the season begins, which will be slightly unusual in that much will be revealed in the first competitive action after a break.
3. Each race week must matter; Google's week begins on a Sunday, and races occur on a Sunday. We create a dummy variable that is 1 for each week that contains a Formula One race.
4. The final week of the season contains much in the way of reflective pieces on the season, and reflections on the winner of the driver's championship. The same is true for the week following the end of the season, presumably including the post-season moves between teams made by drivers before the off season break.
5. Since the year 2009 there has been a recognised mid-season break, which ought to lead to lower search volumes.

The summary statistics of the main variables used in the empirical analysis are presented in Table 1.

As fan interest is a time series, it is vital to take into account persistence across time. This can be thought of as reflecting habitual behaviour by fans interested in Formula One. To do otherwise would leave significant autocorrelation in the residuals. We assume in equation (1) that the error terms are independently and identically normally distributed with constant variance, and hence residual autocorrelation is a violation of this assumption. Hence, some models include lags of y_t in X_t.

The matrix D_t is populated with season (annual) fixed effects, as well as the impulse indicator variables that are selected via indicator saturation.

Then, at the end of the paper, Appendix displays the results of performing similar regression analyses of the Baseline model (3), for both GTW and GTN, where the explanatory variable capturing the certainty of output, *"Winner Known"*, is split out in as many variables as the number of weeks during which the drivers' winner is already known, while the Formula One championship was still on-going. This analysis complements the other results, since it allows us examining in more detail the speed at which people lose interest on the Formula One competition once the winner is known.
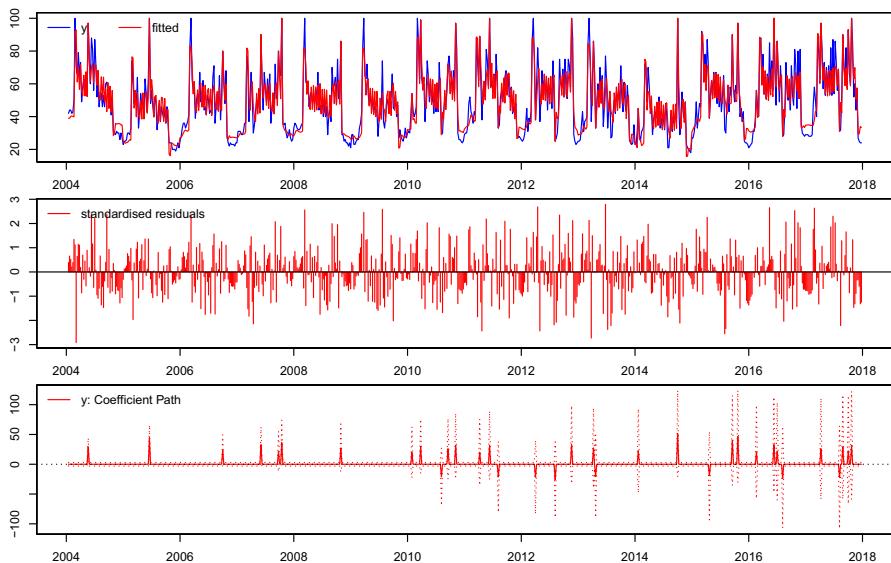
Fig. 3 Model output from column (7), including indicator saturation (IS). Top panel is the actual search volume (GTW) observations (blue line) and the fitted/predicted values from the model (red line). The middle panel is the residuals, standardised using the residual standard error. The bottom panel presents spikes where outliers were detected, with each spike exactly proportional to its size as represented by the coefficient in the finally estimated model

4 Results

This section presents the estimations of using two alternative dependent variables. First, the relative interest on the competition is measured on the basis of the intensity with which Google users search for contents worldwide (GTW), over a 14-year period (from 2004 to 2017). Then, the same analysis is replicated onto a shorter period (from 2008 to 2017), due to data availability, by applying instead an approach based on search intensity of Google news articles (GTN). While the latter approach concerns the relative degree of interest granted by journalist and the media, the former procedure is expected to procure a more comprehensive measurement of the overall degree of interest as shown by sport fans and the general public.

4.1 Analysis with Google Trends Web (GTW)

Initially, the main results are reported in Table 2 for GTW searches. We present a range of models, increasing in the number of explanatory variables that they have. As we proceed from left to right, the models display greater explanatory power and better statistical properties.

We begin in column (1) with a regression of search volume on the variable that captures if the winner is known, while controlling for the increased search volume that corresponds with the actual racing season. Among other findings, these two results stand out: (i) there are many more searches during the season—search volume increases

Table 2 Regression output from regressing "Google Trends Web" (GTW) search volumes on a range of regressors

	Dependent variable: Google Trends Web						
	Races (1)	Season FE (2)	Baseline Model (3)	Raceday (4)	One lag (5)	Two lags (6)	IIS (7)
Constant	34.169***	41.037***	29.583***	29.561***	23.588***	20.730***	25.960***
	(0.978)	(2.193)	(1.469)	(1.331)	(1.560)	(1.637)	(1.602)
Winner known	−8.024***	−10.748***	−10.356***	−12.224***	−9.944***	−10.151***	−8.837***
	(1.680)	(1.747)	(2.174)	(2.358)	(2.187)	(2.216)	(1.490)
Racing season ongoing	22.271***	22.128***	23.494***	13.909***	8.848***	5.614***	6.413***
	(1.236)	(1.208)	(1.118)	(1.061)	(1.237)	(1.305)	(1.010)
Race in that Week				18.673***	18.204***	18.901***	15.218***
				(1.195)	(1.175)	(1.150)	(0.777)
Lag 1 searches					0.209***	0.151***	0.165***
					(0.032)	(0.034)	(0.027)
Lag 2 searches						0.162***	0.173***
						(0.036)	(0.027)
Week before season			53.224***	53.218***	51.182***	50.042***	50.379***
			(1.815)	(1.778)	(1.631)	(1.669)	(1.447)
Week season starts			32.130***	23.036***	17.203***	21.162***	20.983***
			(4.064)	(4.099)	(4.210)	(4.177)	(3.862)

Table 2 continued

	Dependent variable: Google Trends Web						
	Races (1)	Season FE (2)	Baseline Model (3)	Raceday (4)	One lag (5)	Two lags (6)	IIS (7)
Week after season ends			-4.476***	-4.482***	-11.014***	-13.348***	-12.373***
			(1.595)	(1.601)	(1.772)	(2.174)	(1.763)
Second week after season ends			-5.976***	-5.982***	-5.279***	-10.694***	-9.693***
			(1.962)	(1.951)	(1.681)	(2.192)	(1.723)
Season fixed effects	No	Yes	Yes	Yes	Yes	Yes	Yes
Observations	728	728	728	728	727	726	726
R^2	0.303	0.379	0.573	0.722	0.742	0.755	0.856
Adjusted R^2	0.301	0.366	0.562	0.714	0.734	0.747	0.843
Residual Std. Error	15.707	14.963	12.443	10.055	9.695	9.453	7.454
	(df = 725)	(df = 712)	(df = 708)	(df = 707)	(df = 705)	(df = 703)	(df = 667)
F Statistic	157.713***	28.968***	50.000***	91.583***	96.439***	98.518***	68.103***
	(df = 2; 725)	(df = 15; 712)	(df = 19; 708)	(df = 20; 707)	(df = 21; 705)	(df = 22; 703)	(df = 58; 667)

Robust standard errors in parentheses * $p<0.1$; ** $p<0.05$; *** $p<0.01$

by 22 percentage points; and (ii) there is a negative effect on search interest once the winner of the season's championship is known—search volume falls by about 8 percentage points.

The second column adds in annual fixed effects, hence dummies that are one for each calendar year. We do not report these, but they are jointly significant (F statistic 6.7). Their impact is to slightly increase the size of the effect of the winner being known—of uncertainty being taken away—to around 10 percentage points.

In column (3) we add variables that allow for a further impact of the weeks either side of the start and finish of the racing season. The variable for the week before the season begins (the Sunday to Saturday before the race, that includes the practice and qualifying for that race) is very large indeed, suggesting that in that week interest is increased by 53 percentage points. This drops to 28 percentage points in the week that the season has begun. The constant term is 38, giving some sense of the size of these effects. There is also a significant effect at the end of the season, although both negative coefficients are not very large. The effect of the winner of the championship being known falls slightly to just under 10 percentage points.

In the fourth column, we add information on whether a race took place in that particular week. Races during the season tend to be every other week, and sometimes there are longer breaks such as a mid-season break introduced in 2009.[9] If race weeks generate greater interest, then it is important that our model account for this. A race week increases search volume by about 18 percentage points, and has the effect of reducing the size of the racing season coefficient from around 24 to just under 15 percentage points. Search volume is higher in the season, but it is especially high on race weeks. The other notable effect of this is to increase the size of the effect of the winner being known to 12 percentage points.

In columns (5) and (6), we add lagged information. One lag is significant, but insufficient in the sense that a Breusch–Godfrey test for residual autocorrelation rejects the null hypothesis of no autocorrelation. We add two lags in column (6), and both are very significant, reflecting persistence in search patterns. Neither is particularly large, at under 0.2, suggesting that an increase (or decrease) in search activity has a small effect in subsequent weeks. There are three further impacts of doing this: first, the significance of the dummies at the end of the season become larger and significant, at around 10–12 percentage points; second, the effect of the racing season being ongoing falls from 14 to 6 percentage points, though remains very significant; third, the effect of the winner being known falls back to around 10 percentage points.

In column (7), we add in indicators detected using IS to correct for outliers; least squares estimation is not robust to large outliers, and the series presented in Fig. 1 do appear to have a number of spikes and hence potential outliers in them. The detected outliers are very large indeed and are visible in the bottom panel of Fig. 3; in absolute terms, on average they are 27 percentage points. The largest is 51, and the smallest is 17. Of the 36 detected, 28 are positive, suggesting a spike in interest, whereas 8 are small (unexpectedly low levels of search interest), and five of these fall around the time of the mid-season break. It is important to account for outlying observations, both

---

[9] Our *"Race in that week"* variable is 1 for 53% of weeks during the Formula One season, indicating that races are irregular in their frequency during the season.

economically and econometrically. Economically, it is highly likely that one-off events occur that are particularly newsworthy yet bear no resemblance to understanding the impact of the winner being known. It thus makes sense to control for such outlier events. Econometrically, OLS estimation is not robust to outliers, and further, search volumes are scaled by the week with the largest search volume each year. In doing so, some of the coefficients reported reflect this lack of robustness, as they change slightly between columns (6) and (7).

The mean search volume is 26, hence about a quarter of the largest search volume in a given year. A given search week during the racing season shows about 6 percentage points more of search activity, while a week that contains an actual race further increases search by 15 percentage points. The week before the season increases search by 50 percentage points, the first actual week of the season by 21 percentage points, and the last week of the season (so the week that begins with the final race) shows 12 percentage points less search activity, and the week after that 10 percentage points less.

More importantly for the paper's aim, the effect of the winner being known is just under 9 percentage points, which is larger than the effect of the racing season being ongoing, although smaller than the impact of a race taking place in a given week. This suggests that residual, on-going interest in Formula One is mainly driven by races and by uncertainty over who will win the pennant race.

The final model from column (7) is presented in Fig. 3. The top panel presents the actual search volumes over time (a weekly time series) in blue, with the model's fitted (or predicted) values plotted in red. This shows that the main patterns in the data, visible in Fig. 1 above, are largely captured. The middle panel represents the residuals, which are the difference between the actual values and predicted values in the sample. They are scaled by the residual standard error, and hence should be approximately standard normally distributed, exhibiting no autocorrelation, or heteroskedasticity. They all fall within the expected range and indeed show no signs of model misspecification (Ljung–Box AR(1) test p-value 0.16, Ljung–Box ARCH(1) p-value 0.95, Jarque–Bera normality test p-value 0.14).

The bottom panel shows the outliers detected, scaled by the size of their coefficients in the finally estimated model. The majority of them are positive, and large, with a small number of positive ones coming after 2009 when the mid-season break was introduced. (That is, the bottom plot gives some sense of the underlying trends apparent in the data across the series).

4.2 Analysis with Google Trends News (GTN)

In this section, we present the results of the second approach, GTN, which tries to measure the same aspect by relying instead on the amount of times users look for information in "Google News" specifically, rather than just in Google. The structure of Table 3 is the same as Table 2, but for the results of GTN (news articles) search volumes. Again, a range of models is displayed, where additional explanatory variables are included as we progress from left to right. The impact of the winner being known is statistically significant after raceday effects are added to the model in column (4'),

although the coefficient is always smaller than that found for GTW searches. It becomes statistically insignificant for models (5') to (7'), although it is always negative on search volumes.

This difference between GTW and GTN results is intriguing. While the GTW coefficients and significance levels are more intuitive, and fit better with the UO hypothesis, this is not sufficient grounds for accepting them over the GTN results. Figure 4, which plots the actual and fitted values from model (7'), shows markedly different seasonal patterns for GTN search relative to GTW search (Fig. 3), in particular in the periods outside of the regular racing season.

A plausible, although untestable, hypothesis that could explain divergences between the results shown in Table 2 and those of Table 3 may be that established fans will continue to search for news regardless of whether the outcome is known, whereas casual observers keen to try and learn more about the sport may not do so. It suggests that the buzz around a sport that might provoke general searching for information diminishes once the winner is known.

Given that the data for GTW are richer (and thus more reliable) than GTN, and insofar as sponsorship contracts and brands are presumably interested to attract the general public, as potential consumers, and not only Formula One fans, we consider the results obtained for GTW are more relevant—to the paper's scope—than the ones achieved for GTN data.

In the models in the first three columns, there appears to be a large and significant effect of the season being ongoing, of around 12 percentage points. Adding a variable for a race in that particular week cuts the size of the coefficient by a third, and adding in lagged search volumes reduces the coefficient to essentially zero, and insignificant. This suggests that persistence in behaviour, and the existence of race weeks, rather than the season per se, drives news search volumes.

Adding in the extra variables also makes clear that it is the week before the season begins that has the larger effect on news searches rather than the first week itself, the former being twice as large in its effect than the latter.

Figure 4 presents the final column, model (7'), where indicator saturation has been used. As it happened with Fig. 3, these are very large in nature, and primarily positive.

5 Conclusions

This paper investigates one of the most well-known theories in sport economics, the *Uncertainty of Outcome Hypothesis*. That is, that the greater the level of uncertainty, the higher is the level of interest. We consider a case where uncertainty surrounding two different competitive aspects in a sport competition can be identified; the individual event, and the overall pennant race or championship.

Given the astonishing developments in new technologies and their spread-out use worldwide, we advocate that *Google Trends* outcomes provide us with accurate appraisals of the global interest in sport events. The raw data obtained from *Google Trends* are expressed with respect to a reference value of 100, which is attached to the maximum value of the searching period. Then, an index is computed such that all the other figures within the period are re-scaled relative to 100.

Table 3 Regression output from regressing "Google Trends News" (GTN) search volumes on a range of regressors

	Dependent variable: Google Trends News						
	Races	Season FE	Baseline model	Raceday	One lag	Two lags	IIS
	(1')	(2')	(3')	(4')	(5')	(6')	(7')
Constant	29.381***	20.591***	18.878***	18.861***	13.852***	12.454***	10.675***
	(1.518)	(2.686)	(2.087)	(1.819)	(1.902)	(1.849)	(1.463)
Winner known	−13.307***	−4.770***	−3.578**	−5.006**	−3.393	−3.033	−2.252
	(2.874)	(1.610)	(1.575)	(2.211)	(2.078)	(2.120)	(1.860)
Racing season ongoing	12.259***	11.537***	12.185***	4.854***	0.867	−0.770	−0.269
	(1.893)	(1.465)	(1.209)	(1.216)	(1.222)	(1.238)	(0.991)
Race in that week				14.282***	14.116***	14.473***	12.519***
				(1.264)	(1.228)	(1.205)	(0.884)
Lag 1 searches					0.283***	0.226***	0.189***
					(0.042)	(0.043)	(0.031)
Lag 2 searches						0.141***	0.178***
						(0.039)	(0.031)
Week before season			41.338***	41.333***	39.124***	38.501***	39.065***
			(4.949)	(4.968)	(4.870)	(5.033)	(5.060)
Week season starts			38.753***	31.797***	23.909***	26.371***	21.074***
			(6.854)	(6.895)	(6.058)	(6.303)	(4.240)

Table 3 continued

	Dependent variable: Google Trends News						
	Races (1')	Season FE (2')	Baseline model (3')	Raceday (4')	One lag (5')	Two lags (6')	IIS (7')
Week after season ends			−5.462**	−5.467**	−11.744***	−12.704***	−11.827***
			(2.647)	(2.631)	(3.128)	(3.234)	(3.138)
Second week after season ends			−7.562**	−7.567**	−6.385***	−9.883***	−10.246***
			(2.980)	(2.949)	(2.369)	(2.865)	(2.998)
Season fixed effects	No	Yes	Yes	Yes	Yes	Yes	Yes
Observations	520	520	520	520	519	518	518
R^2	0.077	0.477	0.617	0.694	0.730	0.737	0.868
Adjusted R^2	0.074	0.466	0.605	0.685	0.721	0.728	0.856
Residual Std. Error	20.239	15.367	13.208	11.809	11.116	10.969	7.974
	(df = 517)	(df = 508)	(df = 504)	(df = 503)	(df = 501)	(df = 499)	(df = 474)
F Statistic	21.632***	42.164***	54.098***	71.414***	79.580***	77.875***	72.620***
	(df = 2; 517)	(df = 11; 508)	(df = 15; 504)	(df = 16; 503)	(df = 17; 501)	(df = 18; 499)	(df = 43; 474)

Robust standard errors in parentheses *p<0.1; **p<0.05; ***p<0.01
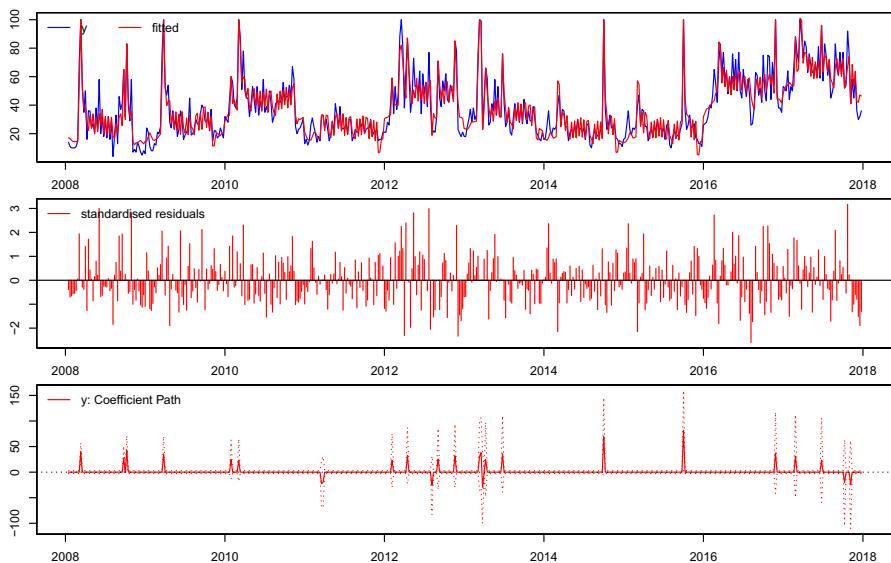
Fig. 4 Model output from column (7'), including indicator saturation. Top panel is the actual GTN search volume observations (blue line) and the fitted/predicted values from the model (red line). The middle panel is the residuals, standardised using the residual standard error. The bottom panel presents spikes where outliers were detected, with each spike exactly proportional to its size as represented by the coefficient in the finally estimated model

We examine whether knowledge about the overall outcome of a pennant race will affect the interest in the individual events that contribute towards that pennant race, by looking at Formula One races and the overall driver's championship. Often towards the end of a season, it is known already which driver has won the driver's championship, and hence, from this point on, only individual race-level uncertainty remains.

In the paper, we describe how, along with the volume of general searches in websites, a double-check more robust analysis can be performed by examining also figures on the searching intensity of news articles, which are also provided by *Google Trends*.

The results of our empirical analysis are not totally conclusive, deserving further research effort. Concerning the degree of general interest manifested by the general public, as captured by the relative intensity in searching activity (GTW), once pennant-race uncertainty is removed, we find that search interest declines significantly, and by a large amount—about 10 percentage points, when average search activity is at a level of about 25. That is, about a third of search interest is lost. Some interest remains, reflecting that each individual race is interesting and contains within it uncertainty. Nevertheless, if we adopt the alternative approach of measuring the people degree of interest as manifested by their search intensity of news articles on the matter (GTN), it is more dubious to conclude that getting to know who is the driver's champion of the year significantly reduces the interest on the Formula One competition.

Given that we have richer data on the general Google searches (GTW), and because sponsorship contracts and brand development do presumably want to attract the general public as potential customers (rather than just Formula One fans), we tend to believe

that the results from the first analysis (GTW) are more reliable than those derived from the second approach (GTN).

These results are of interest to the designers of sporting competitions; different structures in terms of the number of points available in individual events contributing to a pennant race, can affect how quickly the overall winner is determined. Search volume can be monetised, and as such we quantify that around a third of such potential activity is lost once the pennant race ceases to be uncertain.

Among other policy implications, our findings suggest that—to prevent a loss of interest in sport events—the way how sport tournaments' are designed should be revised. Moreover, as concern the organisers of sport championships, they may want to alter the competition structure in such a way that the drivers' winner would not be known until very late in the season. Otherwise, we estimate that around one third of the interest in sport competitions (and the corresponding potential revenues) may be lost.

In the near future, we plan replicating a similar analysis to team-sport leagues, particularly the "Big-five" domestic Football leagues in Europe. Further research is needed also to examine other sport disciplines and to explore additional aspects related to outcome uncertainty other than the approach focussed on the certainty about the final winner.

Appendix

Appendix shows the results of replicating the estimations for baseline models (3) and (3'), but where the output certainty (captured by the *"Winner Known"* variable), rather than a single variable, has been split in as many variables as the number of weeks in which the drivers' winner was already known. (The Formula One championship was still on-going. This figure reached even 8 weeks in one of the years under examination.) This new approach allow us examining how fast the interest drops once the Drivers' winner is known and to verify if there may be a rebound effect as time (weeks) go by.

The estimated models are equivalent to model (3) in Table 2, and model (3') in Table 3. These models correspond to estimations without introducing 'Indicator Saturation" techniques, and after having replaced the variable "certainty of output" by

Table 4 Regression output-baseline model for the explanatory variable "WinnerKnown" split out into the number of weeks

	Dependent variable:							
	GTrends WEB Model (3)	t-stat	GTrends WEB Model (3A)	t-stat	GTrends NEWS Model (3')	t-stat	GTrends NEWS Model (3A')	t-stat
Constant	39.7065 ***	(18.03)	39.7385 ***	(18.12)	54.3241 ***	(24.82)	54.3882 ***	(24.89)
Winner known	− 9.8228 ***	(− 5.82)			− 3.5775 *	(− 2.27)		
Racing season ongoing	23.9710 ***	(26.98)	24.1387 ***	(26.88)	12.1848 ***	(10.07)	12.3031 ***	(10.08)
Week before season	53.3458 ***	(34.33)	53.3461 ***	(34.45)	41.3375 ***	(8.35)	41.3388 ***	(8.29)
Week season starts	28.7319 ***	(7.34)	28.5645 ***	(7.23)	38.7527 ***	(5.65)	38.6357 ***	(5.60)
Week after season ends	− 3.2256 *	(− 2.35)	− 3.2252 *	(− 2.30)	− 5.4624 *	(− 2.06)	− 5.4611 *	(− 2.05)
2nd week after season ends	− 4.7256 **	(− 2.93)	− 4.7252 **	(− 2.89)	− 7.5624 *	(− 2.54)	− 7.5611 *	(− 2.53)
Winner known_1			− 10.1795 ***	(− 3.50)			− 7.5278	(− 1.76)
Winner known_2			− 7.1795 **	(− 3.24)			− 3.5278	(− 1.32)
Winner known_3			− 4.8983	(− 1.50)			0.4287	(0.22)
Winner known_4			− 13.2289 ***	(− 3.88)			− 5.5712 *	(− 2.24)
Winner known_5			− 13.2886 ***	(− 8.47)			− 7.7339 **	(− 3.43)
Winner known_6			− 15.0313 ***	(− 4.51)			− 5.1366 ***	(− 4.05)
Winner known_7			− 17.0313 ***	(− 4.62)			− 3.1366 *	(− 2.47)
Winner known_8			− 12.7916 ***	(− 7.43)				
SeasonDummies	Yes		Yes		Yes		Yes	
Observations	728		728		520		520	
R^2	0.5730		0.5773		0.6169		0.6185	

Robust standard errors in parentheses *$p<0.1$; **$p<0.05$; ***$p<0.01$

up to 8 variables, to capture if there are discrepancies of the decline in interest over the weeks. For the sake of robustness, the estimations shown in the table involve the results of both the analysis based on the intensity with which Google users search for both news articles (GTN) and for general web contents (GTW).

$$y_t = \beta_0 + \gamma \cdot D_t + \phi \cdot X_t + \sum_{i=1}^{n} \beta_i \cdot Winner\,Known_t + \mu_t, \qquad \varepsilon_t \sim N(0, \sigma^2)$$

$$(2)$$

The results show that getting to know the driver's winner has a negative effect on fans' degree of interest in the Formula One championship, which is still statistically significant even after 8 weeks of certainty, even if the drop of interest (as captured by GTW) is small after 1, 2 and specially 3 weeks. Furthermore, if we relay on news, rather than general searches in Google, the reduction in followers interest only appears to be significant starting in the fourth week of certainty.

References

Aguiar-Noury A, Garcia-del-Barrio P (2019) Global brands in soccer: Identifying low-risk projects and business opportunities. J Entrep Pub Policy 8(1):62–83

Aguiar-Noury A, Garcia-del-Barrio P (2021) Global hierarchy of team-sport leagues based on internet searches and revenues: Europe versus America. Eur J Int Manag 15(2/3):283–313

Bai J, Perron P (2003) Computation and analysis of multiple structural change models. J Appl Economet 18(1):1–22

Borland J, MacDonald R (2003) Demand for sport. Oxf Rev Econ Policy 19(4):478–502

Bulut L (2018) Google trends and the forecasting performance of exchange rate models. J Forecast 37(3):303–315

Buraimo B, Forrest D, McHale IG, Tena JD (2020) Unscripted drama: Soccer audience response to suspense, surprise, and shock. Econ Inq 58(2):881–896

Cairns J, Jennett N, Sloane PJ (1986) The economics of professional team sports: a survey of theory and evidence. J Econ Stud 13(1):3–80

Castle, J.L.; Doornik, J.A. and Hendry, D.F. (2008). Model selection when there are multiple breaks. Working paper

Choi H, Varian H (2012) Predicting the present with google trends. Econ Rec 88(s1):2–9. https://doi.org/10.1111/j.1475-4932.2012.00809.x

Coates D, Humphreys BR (2010) Week to week attendance and competitive balance in the National Football League. Int J Sport Financ 5(4):239–252

Coates D, Humphreys BR (2012) Game attendance and outcome uncertainty in the National Hockey League. J Sports Econ 13(4):364–377

Czarnitzki D, Stadtmann G (2002) Uncertainty of outcome versus reputation: empirical evidence for the first german football division. Empir Econ 27(1):101–112. https://doi.org/10.1007/s181-002-8361-7

Ely J, Frankel A, Kamenica E (2015) Suspense and surprise. J Polit Econ 123(1):215–260

Fort R, Quirk J (1995) Cross-subsidization, incentives, and outcomes in professional team sports leagues. J Econ Lit 33:1265–1299

García J, Rodríguez P (2002) The determinants of football match attendance revisited: Empirical evidence from the Spanish football league. J Sports Econ 3(1):18–38

Garcia-del-Barrio P (2018) Media value methodology and global sport industries: football versus Formula One. Int J Sport Manag Mark 18(3):241–266

Garcia-del-Barrio P, Gomez-Gonzalez C, Sánchez-Santos JM (2020) Popularity and visibility appraisals for computing olympic medal rankings. Soc Sci Q 101(5):2137–2157

Hendry DF, Doornik JA and Pretis F (2013) Step-indicator saturation. Economics Series Working Papers 658, University of Oxford, Department of Economics

Humphreys BR (2002) Alternative measures of competitive balance in sports leagues. J Sports Econ 3(2):133–148

Humphreys BR, Zhou L (2015) The Louis-Schmelling paradox and the league standing effect reconsidered. J Sports Econ 16(8):835–852

Jennett N (1984) Attendances, uncertainty of outcome and policy in Scottish league football. Scottish J Political Econ 31(1):176–198

Judde C, Booth R, Brooks R (2013) Second place is first of the losers: an analysis of competitive balance in formula one. J Sports Econ 14(4):411–439

Késenne S (2014) The economic theory of professional team sports: An analytical treatment. (Section: 2.2.1: Uncertainty of outcome and competitive balance. and following pages). Edward Elgar Publishing, p 11

Késenne S (2000) Revenue sharing and competitive balance in professional team sports. J Sports Econ 1(1):56–65

Kringstad M, Gerrard B (2007) Beyond competitive balance. In: Parent MM, Slack T (eds) International perspectives on the management of sports. Elsevier Academic Press, Burlington, MA

Lee YH, Fort R (2008) Attendance and the uncertainty-of-outcome hypothesis in baseball. Rev Ind Organ 33(4):281–295. https://doi.org/10.1007/s11151-008-9193-9

Manasis V, Avgerinou V, Ntzoufras I, Reade JJ (2013) Quantification of competitive balance in European football: development of specially designed indices. IMA J Manag Math 24(3):363–375

Manasis V, Ntzoufras I and Reade JJ (2015) Measuring competitive balance and uncertainty of outcome hypothesis in European football. arXiv preprint arXiv: 1507.00634

Mastromarco C, Runkel M (2009) Rule changes and competitive balance in formula one motor racing. Appl Econ 41(22–24):3003–3014

Neale P (1964) The peculiar economics of professional sports: a contribution to the theory of the firm in sporting competition and in market competition. Quart J Econ 78(1):1–14

Owen, D. (2014). Measurement of competitive balance and uncertainty of outcome. Chapter 3 in: Handbook on the Economics of Professional Football (Edited by Goddard, John and Peter J. Sloane). Edward Elgar. Cheltenham, UK - Norhampton, MA, USA

Pawlowski T, Anders C (2012) Stadium attendance in German professional football-The (un) importance of uncertainty of outcome reconsidered. Appl Econ Lett 19(16):1553–1556

Pawlowski T, Nalbantis G, Coates D (2018) Perceived game uncertainty, suspense and the demand for sport. Econ Inq 56(1):173–192

Peel DA, Thomas DA (1992) The demand for football: Some evidence on outcome uncertainty. Empir Econ 17:323–331

Perez L, Puente V, Rodriguez P (2017) Factors determining TV soccer viewing: Does uncertainty of outcome really matter? Int J Sport Financ 12(2):124–139

Pretis F, Reade JJ and Sucarrat G (2017) Gets: General-to-specific (GETS) modelling and indicator saturation methods, 2017. URL http://CRAN.R-project.org/package=gets. R package version 0.12

Rottenberg S (1956) The baseball players labor market. J Polit Econ 64(3):242–58

Sanderson AR (2002) The many dimensions of competitive balance. J Sports Econ 3(2):204–228

Szymanski S (2001) Income inequality, competitive balance and the attractiveness of team sports: some evidence and a natural experiment from English soccer. Econ J 111(469):F69–F84

Szymanski S (2003) The economic design of sporting contests. J Econ Lit 41:1137–87

Whitney JD (1988) Winning games versus winning championships: the economics of fan interest and team performance. Econ Inq 26(4):703–724

Zimbalist AS (2002) Competitive balance in sports leagues: an introduction. J Sports Econ 3(2):111–121