Does Match Uncertainty Increase Attendance? A Non-Regression Approach

Lahvicka, Jiri

23 July 2013

Does Match Uncertainty Increase Attendance? A Non-Regression Approach

Jiří Lahvička

E-mail: jirka@lahvicka.cz

The uncertainty of outcome hypothesis predicts that more balanced sports matches should attract higher attendances, but the empirical evidence is mixed at best. First, this article shows that the inconsistent findings in the literature could be explained by wrongly specified regressions. Second, a new approach to analyzing the effect of match uncertainty is proposed. Using data about nine seasons of the English Championship, the article shows that in a pair of matches where both home teams are slight favorites, a switch of the corresponding away teams would decrease the total attendance by several percent, while the opposite is true if both home teams are underdogs or strong favorites. These results suggest that attendance demand is a bell-shaped function of match balance that is maximized if teams of the same quality play against each other.

1 Introduction

Do more balanced sports matches attract higher attendances? The uncertainty of outcome hypothesis (Rottenberg, 1956; Neale, 1964) certainly predicts so, but the empirical evidence is mixed at best. So far, the link between match uncertainty and attendance has been examined by regressing individual match attendance (or its logarithm) on variables representing qualities of both teams, other variables influencing attendance (ticket price, team rivalry, distance between teams, weather…), and a variable measuring how the match is balanced. 18 such studies reviewed in Borland and McDonald (2003) investigated different sports (mostly soccer and baseball), used different ways of measuring team quality (team ranks or points/goals per game) and match uncertainty (difference in team ranks or points per game; absolute value of betting spread; quadratic specification of home win probability derived from betting odds), and arrived at different results; some studies found that higher match uncertainty increases attendance, some found the opposite, some found that attendance increases with home win probability (and possibly starts decreasing if home win probability is higher than 0.6-0.7), some found no significant effect.

Similarly contradictory results can also be found in more recent research. Buraimo and Simmons (2008) modeled English Premier League attendance and concluded that attendance is minimized if home win probability derived from betting odds equals about 0.35. Buraimo and Simmons (2009) obtained a similar result for Spanish soccer. However, Benz et al. (2009) found that for one model specification, German Bundesliga attendance (excluding season tickets) was maximized for home team win probability equal to 0.53. Contradictory results for German Bundesliga were obtained by Pawlowski and Anders (2012); in one regression specification, attendance decreased if home team was a favorite rather than outsider; in another specification; higher match uncertainty decreased attendance. Coates and Humphreys (2011) claimed that the previous inconsistent results were due to linear or quadratic specifications of match uncertainty; their results for the NHL indicate that the attendance increases if the home team is a strong favorite or a slight underdog.

This article makes two contributions. First, three simple simulated data sets with no impact of match uncertainty on attendance are used to show that different commonly used regression specifications produce different (and wrong) results about the link between match uncertainty and attendance. This could explain the inconsistent findings in the literature, especially if the actual impact of match uncertainty is weak or nonexistent. Second, a new approach to analyzing the effect of match uncertainty on attendance is proposed. Using data about nine seasons of the English Championship, the article shows that in a pair of matches where both home teams are slight favorites, a switch of the corresponding away teams would decrease the total attendance. On the other hand, if both home teams are underdogs or strong favorites, switching the away teams would increase the total attendance. However, the magnitude of such attendance changes is quite small (several percent). These results are consistent with the uncertainty of outcome hypothesis and suggest that attendance demand is a bell-shaped function of match balance that is maximized if teams of the same quality play against each other.

2 Data

The proposed method of measuring the impact of match uncertainty on attendance is demonstrated on the data set consisting of nine regular seasons (2004/05-2012/13) of the second-highest English soccer league; English Championship. This competition still attracts a lot of spectators, but attendances only rarely come close to the stadium capacity, so the attendance demand for each match is directly observable.

In each season of the Championship, 24 teams play one home and one away match against each other, so there are 552 matches in each season and 4,968 matches in the whole Championship data set. The relevant data for each match are its attendance, which was downloaded from the website worldfootball.net, and the corresponding betting odds, which were obtained from the website football-data.co.uk.[1] The betting odds were converted in a standard way to home win, draw and away win probabilities and these probabilities were averaged across different bookmakers.[2]

For each match, match balance was calculated as the home win probability plus one half of the draw probability; this variable is similar to home win probability used in many previous articles, but has the advantage that it is exactly 0.5 for perfectly balanced matches with each team having the same probability of winning. The descriptive statistics for variables *Attendance* and *MatchBalance* are provided in Table 1.

	Average	Min	Percentiles					Max
			0.1	0.25	0.5	0.75	0.9	
Attendance	17,632	1,211	9,492	12,822	17,219	22,267	25,652	52,181
MatchBalance	0.5743	0.2641	0.4758	0.5182	0.5699	0.6340	0.6806	0.8083

Table 1: Descriptive statistics of the Championship data set, N = 4,968

The average match balance value is above 0.5; this reflects the home team advantage. Most match balance values (80%) are concentrated in the 0.48-0.68 interval, which means that it is hard to say much about what happens to match attendance outside this interval.

---

[1] The attendance data were downloaded on July 7th, 2013. One missing attendance figure was obtained from the website www.11v11.com. The betting odds on home win, draw, and away win were downloaded on June 10th, 2013, and provided by major bookmakers William Hill, Bet&Win, and Interwetten. Although some betting odds were missing, there was at least one set of betting odds for each match.
[2] To convert betting odds into probabilities, they are first inverted. The sum of these inverted numbers is more than one to allow for bookmaker's profit, so the inverted numbers are divided by this sum to obtain the home win, draw, and away win probabilities.

3 The problem with regression

This section shows that commonly used regression specifications can easily produce misleading results about the relationship between match uncertainty and attendance. This is demonstrated on three simple simulated data sets with no impact of match uncertainty on attendance.

To construct each data set, let's assume there are 24 teams in a competition (the same as in the Championship data set) and team qualities are uniformly distributed on the interval [0; 1]. This means that team i's quality $Quality_i = (i – 1)/23$; the $Quality$ variable corresponds to the normalized rank or points per game used in other studies. The teams play one home and one away match against each other, generating one complete season of 552 matches. Let's further assume that each team attracts a fixed number of spectators to its home matches and that there is a different (smaller) fixed number of spectators that travel with the team to its away matches. Both numbers are increasing functions of team quality.[3]

There is no special reason why the relationship between team quality and the number of spectators should be linear; in fact, a non-linear relationship between team rank and points per game guarantees that it is not the case in at least some previous studies. The data sets cover the three simplest cases: attendance is an exactly linear function of quality (Data set 1), attendance is a concave function of quality (Data set 2), and attendance is a convex function of quality (Data set 3).

To produce plausible total attendance numbers (similar to the Championship data set), the home and away spectator numbers $Home_i$ and $Away_i$ attracted by team i are set equal to the following expressions:

Data set 1 (linear): $Home_i = 5{,}000 + 25{,}000 * Quality_i$, $Away_i = 5{,}000 * Quality_i$

Data set 2 (concave): $Home_i = 5{,}000 + 25{,}000 * Quality_i^{0.8}$, $Away_i = 5{,}000 * Quality_i^{0.8}$

Data set 1 (convex): $Home_i = 5{,}000 + 25{,}000 * Quality_i^{1.25}$, $Away_i = 5{,}000 * Quality_i^{1.25}$

The attendance of a match between teams i and j ($Attendance_{ij}$) is simply the sum of spectators attracted by the home team ($Home_i$) and spectators travelling with the away team ($Away_j$):

Data set 1 (linear): $Attendance_{ij} = 5{,}000 + 25{,}000 * Quality_i + 5{,}000 * Quality_j$

Data set 2 (concave): $Attendance_{ij} = 5{,}000 + 25{,}000 * Quality_i^{0.8} + 5{,}000 * Quality_j^{0.8}$

Data set 1 (convex): $Attendance_{ij} = 5{,}000 + 25{,}000 * Quality_i^{1.25} + 5{,}000 * Quality_j^{1.25}$

As assumed above, fans care only about the quality of their own team, so attendance in the simulated data sets does not depend on match uncertainty at all. This is also clear from the

---

[3] A similar assumption was made in Peel and Thomas (1992). Their estimation results also show that match attendance increases with qualities of both teams (measured by team ranks) with the home team's quality having a stronger influence.

attendance formulas – they are additively separable (there is no interaction between qualities of both teams). Therefore, any valid method of measuring the impact of match uncertainty on attendance should conclude that the impact is zero. To test whether this is true for common regression specifications, a variable $MatchBalance_{ij}$ is defined in the following way for each match using the logistic function:

$MatchBalance_{ij} = 1/(1 + exp(Quality_j – Quality_i – 0.25))$

The match balance variable is an increasing function of home team's quality and a decreasing function of away team's quality, so a higher value indicates that the home team is more likely to win (the number 0.25 provides the home advantage). The match balance values in each simulated data set range from 0.32 to 0.78 with the average of 0.56, closely mimicking the match balance values in the Championship data set that were calculated as the home win probability plus one half of the draw probability.

An additional variable $MatchUncertainty_{ij}$ measures how close a specific match is to the ideal balance of 0.5:

$MatchUncertainty_{ij} = 1 – 2 * |MatchBalance_{ij} – 0.5|$

If a match is perfectly balanced, $MatchUncertainty_{ij}$ equals 1; on the other hand, if one team is sure to win, $MatchUncertainty_{ij}$ goes down to 0. This variable is analogical to variables such as the difference in team ranks, difference in points per game (possibly adjusted for home team advantage), or absolute value of betting spread used in other studies.

A researcher wanting to use a regression approach to measure the impact of match uncertainty on attendance could choose from various regression specifications. Probably the simplest one copies the attendance-generating formula of Data set 1 and adds the $MatchUncertainty$ variable:

$Attendance_{ij} = \beta_0 + \beta_1 * Quality_i + \beta_2 * Quality_j + \beta_3 * MatchUncertainty_{ij} + \varepsilon$

This simple regression specification can be modified by replacing attendance with its logarithm (since many variables are expected to influence attendance by a given percentage instead of by a given number of spectators); by replacing home team quality with a set of dummies for each home team (home fixed effects); by also replacing away team quality with a set of dummies (all fixed effects); or by replacing the $MatchUncertainty$ variable with a quadratic specification of the $MatchBalance$ variable (i.e. $\beta_3 * MatchBalance_{ij} + \beta_4 * MatchBalance_{ij}^2$). The estimated effects of match uncertainty on attendance for all three simulated data sets using twelve possible regression specifications are summarized in Table 2.[4]

---

[4] All models were estimated with heteroskedasticity-consistent standard errors. Results are reported as significant if p-value < 0.05.

	Data set 1 (linear)	Data set 2 (concave)	Data set 3 (convex)
Attendance No fixed effects Uncertainty	Zero effect	Higher uncertainty increases attendance	Higher uncertainty decreases attendance
Log of attendance No fixed effects Uncertainty	Insignificant	Higher uncertainty increases attendance	Higher uncertainty decreases attendance
Attendance Home fixed effects Uncertainty	Zero effect	Higher uncertainty increases attendance	Higher uncertainty decreases attendance
Log of attendance Home fixed effects Uncertainty	Higher uncertainty decreases attendance	Higher uncertainty decreases attendance	Higher uncertainty decreases attendance
Attendance All fixed effects Uncertainty	Zero effect	Zero effect	Zero effect
Log of attendance All fixed effects Uncertainty	Higher uncertainty decreases attendance	Higher uncertainty decreases attendance	Higher uncertainty decreases attendance
Attendance No fixed effects Quadratic balance	Zero effect	Attendance maximized if balance = 0.41	Attendance minimized if balance = 0.54
Log of attendance No fixed effects Quadratic balance	Insignificant	Insignificant	Attendance minimized if balance = 0.15
Attendance Home fixed effects Quadratic balance	Zero effect	Attendance maximized if balance = 1.30	Attendance minimized if balance = 1.01
Log of attendance Home fixed effects Quadratic balance	Attendance minimized if balance = 0.18	Attendance minimized if balance = 0.11	Attendance minimized if balance = 0.23
Attendance All fixed effects Quadratic balance	Zero effect	Zero effect	Zero effect
Log of attendance All fixed effects Quadratic balance	Attendance minimized if balance = 0.25	Attendance minimized if balance = 0.25	Attendance minimized if balance = 0.25

Table 2: Estimated effect of match uncertainty on attendance for simulated data sets

Since there is no actual relationship between match uncertainty and attendance in the simulated data sets, most regression results are wrong. The wrong results are caused by a not exactly linear relationship between attendance (or its logarithm) and team quality variables; fitting a linear model leads to a specific pattern of residuals that is then captured by match uncertainty or balance variables (which are themselves determined by team qualities). The only specification that produces correct results for all three data sets includes fixed effects for both home and away teams and a non-logarithmic attendance. However, even this specification could be problematic, since it must impose a specific functional form on the relationship between match balance and attendance. There is no theoretical reason why this relationship should be linear or quadratic – Coates and Humphreys (2011) hypothesized an asymmetric relationship (fans preferring matches where home teams are strong favorites or slight underdogs), but it could easily be S-shaped (most fans do not care about match uncertainty, but some fans will attend a match only if the home team is sufficiently favored) or bell-shaped (most fans do not care about match uncertainty, but some fans will attend a match only if it is balanced enough). Again, an incorrect specification would tend to provide misleading results.

Regression results are not only suspect, but also hard to interpret; they do not say how attendance would change if match uncertainty changed and all the other variables stayed constant, since match uncertainty cannot change without also changing team qualities. Some authors address this by taking the estimated attendance demand function and asking what would happen if the league structure changed (Dobson et al., 2001) or if teams were more evenly balanced (Forrest and Simmons, 2002; Buraimo and Simmons, 2009). However, if the estimated demand function is misspecified, this approach could still lead to incorrect conclusions.

As shown above, different regression specifications lead to different (and mostly wrong) conclusions about the relationship between match uncertainty and attendance. The range of results reported in Table 2 is in fact similar to those found in the literature for real data. Therefore, the inconsistent findings in the literature could easily be caused by misspecified regressions, especially if the actual impact of match uncertainty on attendance is weak or nonexistent. Clearly, a different approach is necessary. Such an approach should fulfill three criteria; first, it should not find any effect of match uncertainty on attendance in any of the simulated data sets; second, it should not assume any specific functional form of the match uncertainty-attendance relationship; third, its results should be easy to interpret. Exactly such an approach is described in the next section.

4 A non-regression approach

This section presents an approach to examining the link between match uncertainty and attendance that does not use a regression. The proposed method is demonstrated on the Championship data set described above. The main idea is that although it is not possible to change uncertainty of a match between two fixed teams, it is possible to change uncertainties of matches between two fixed sets of

teams by pairing them in different ways and then analyze what happens to the total attendance of such match combinations.

The approach (as applied to the Championship data set) starts with all combinations of two home teams (H_1, H_2) and two away teams (A_1, A_2) in a given season. Since there are 24 teams in a season, there are 24 * 23/2 possible home team pairs and 22 * 21/2 possible away team pairs, giving 24 * 23 * 22 * 21/4 = 63,756 combinations per season and 63,756 * 9 = 573,804 combinations for the whole nine-season data set. Each combination of two home teams and two away teams can be matched in two different ways: H_1-A_1 + H_2-A_2 or H_1-A_2 + H_2-A_1. Those two possible pairs of matches will be different in terms of both match balance and total attendance. To test whether some specific value of match balance (*IdealBalance*) could maximize (or minimize) match attendance, only those combinations of home and away teams are selected where balances of both matches in one pair are close to *IdealBalance* (balanced pair), while one match balance in the second pair is much lower than *IdealBalance* and the other is much higher (unbalanced pair). More formally, the conditions for not discarding a combination of teams are the following (α and β are parameters, $\beta \geq \alpha > 0$):

Balanced pair: *IdealBalance – α ≤ Balances of both matches ≤ IdealBalance + α*

Unbalanced pair: *One match balance < IdealBalance – β < IdealBalance + β < The other match balance*

Decreasing α and increasing β creates a bigger contrast between the pairs of matches, but decreases the number of team combinations that are not discarded, so the exact values should be chosen depending on the size of the data set. The specific values of α and β chosen for the Championship data set are $\alpha = 0.03$ and $\beta = 0.09$.

After keeping only the team combinations where one pair of matches between them is sufficiently balanced and the other is sufficiently unbalanced, the total attendance of each pair of matches for each such combination is calculated using the actual historical attendances. If both pairs of matches for the same team combination have exactly the same attendance, they are discarded (extremely rare). Finally, the probability that the more balanced combination of matches will also be more attended is estimated by the relative frequency of this happening among all non-discarded team combinations. Of course, this probability by itself does not say much about the size of the effect (it depends on how strong the other factors that influence the attendance are). Therefore, the last value to be calculated is the ratio of the total attendance of all balanced pairs to the total attendance of all unbalanced pairs.

The method might be further illustrated with the following example of testing whether perfectly balanced matches (*IdealBalance = 0.50*) might maximize attendance: One possible combination of teams in the 2012/13 season is Middlesbrough (16) and Bristol (24) as home teams and Leicester (6) and Blackburn (17) as away teams (the numbers in parentheses are their final ranks in the season). All relevant match balances and actual attendances are provided in Table 3.

	Match	Match balance	Attendance	Total attendance
Balanced pair	Middlesbrough (16) – Leicester (6)	0.5046	15,679	30,336
	Bristol (24) – Blackburn (17)	0.4872	14,657	
Unbalanced pair	Middlesbrough (16) – Blackburn (17)	0.6080	22,882	35,960
	Bristol (24) – Leicester (6)	0.3324	13,078	

Table 3: Example of match balances and attendances in two possible match pairs

The teams can be paired in two different ways. Both Middlesbrough – Leicester and Bristol – Blackburn matches have their balances between 0.47 and 0.53, so they are a balanced pair. After switching away teams, the balance of one match (Middlesbrough – Blackburn) is greater than 0.59, while the balance of the other match (Bristol – Leicester) is lower than 0.41, so they are an unbalanced pair. The higher total attendance of the unbalanced pair (35,960 instead of 30,336 spectators) suggests that perfectly balanced matches are not ideal for maximizing attendance. This conclusion is confirmed by going through all 2,734 team combinations where one pair of matches is balanced, the other unbalanced, and total attendances of these pairs are different; only in 44.7% of such cases, the balanced pair is more attended. However, the actual attendance difference is not big; the attendance of the balanced pair is on average 98.2% of the attendance of the unbalanced pair.
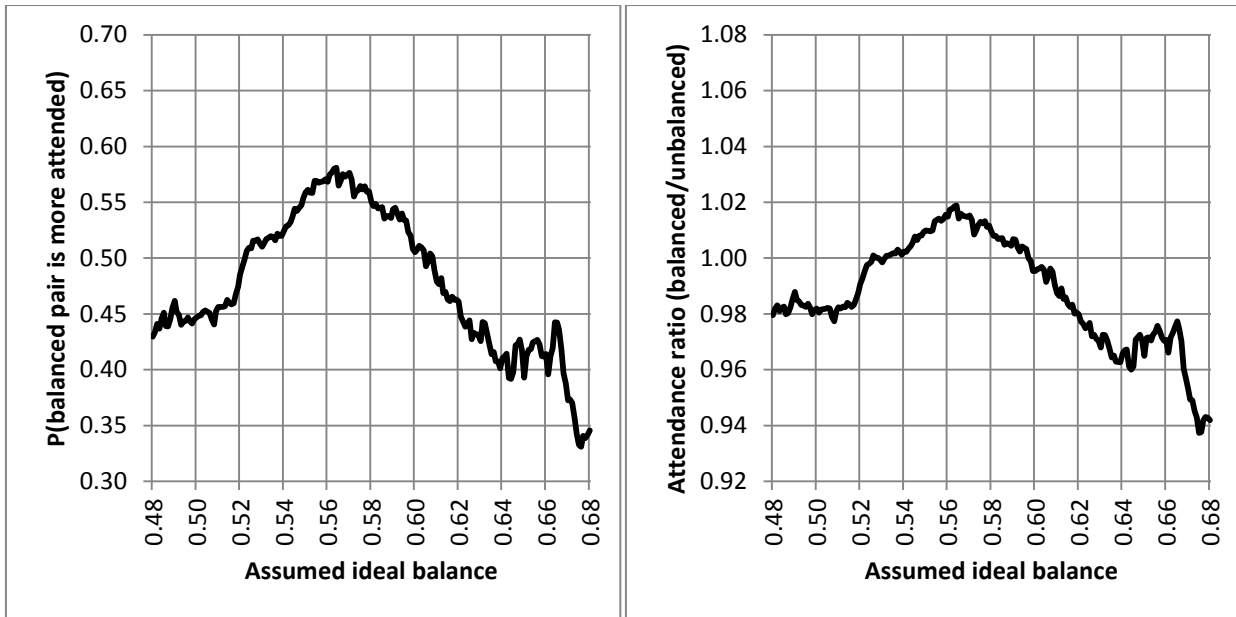

Figure 1: Test results for various values of assumed ideal balance

8

Figure 1 provides test results for all values of *IdealBalance* between 0.48 and 0.68.[5] The left panel shows the estimated probability that the balanced pair of matches will be more attended, while the right panel shows the attendance ratio between all balanced and all unbalanced pairs. For *IdealBalance* values between 0.52 – 0.60, the probability is above 0.5 and the attendance ratio is more than 1, while for *IdealBalance* values outside this interval, the probability is below 0.5 and the attendance ratio is less than 1. Both the probability and the attendance ratio are maximized for *IdealBalance* about 0.56, which is close to the median match balance value in the data set, i.e. typical for two teams of roughly the same quality with home team having a better chance of winning because of home advantage.

The results are robust to changing *α* and *β* (e. g. *α = 0.04, β = 0.06*), discarding team combinations where the change in total attendance between match pairs is too low (e.g. less than 10%), discarding team combinations where the difference between attendances of balanced pair matches is too high (e.g. more than 10 or 20%), or restricting the analysis to a subset of seasons.

How to interpret these findings? First, the results indicate that match uncertainty indeed influences attendance; if there were no link between match uncertainty and attendance, switching away teams would have no systematic effect on total attendance, the probability that the balanced pair is more attended would always be 0.5, and the attendance ratio would always be 1.[6] Indeed, for all three simulated data sets introduced in the previous section (or any other additively separable attendance demand function with no error term), switching away teams would keep the total attendance exactly the same.

Second, if both matches in a pair have approximately the same balance that is between 0.52 – 0.60 (i.e. both home teams are slight favorites), switching away teams that would unbalance both matches would likely decrease the total attendance. However, the attendance change would be small (2% on average if both original match balances are close to 0.56). On the other hand, if both matches in a pair have approximately the same balance that lies outside the 0.52 – 0.60 interval (i.e. both home teams are underdogs or strong favorites), an away-team switch that would unbalance both matches would increase the total attendance. Again, the average attendance change would be just several percent.

Third, the results can be used to infer the shape of the demand function showing how attendance depends on match balance. For *IdealBalance* values between 0.52 and 0.60, a balanced pair of matches has a higher attendance than an unbalanced pair, suggesting a concave shape of the demand function in this region.[7] On the other hand, for *IdealBalance* outside this interval, a balanced

---

[5] The interval ranges from the 10th to the 90th percentile of match balance values in the Championship data set. The test was done for all *IdealBalance* values that are multiples of 0.001, i.e. 0.480, 0.481, 0.482 … 0,679, 0.680. The number of usable team combinations for each *IdealBalance* value is always more than 500 (more than 3,000 on average).

[6] Of course, the non-regression method just estimates the probability that the balanced pair is more attended, so these estimates would not be exactly 0.5 even with zero impact of match uncertainty. However, the null hypothesis of zero impact of match uncertainty can be rejected by a test described in Appendix.

[7] For this conclusion to hold, the average match balance of the balanced pair should be the same as the average balance of the unbalanced pair. This is approximately true in the data.

pair of matches has a lower attendance than an unbalanced pair, so the demand function should be convex. A plausible attendance demand function consistent with these results would be bell-shaped, maximized at match balance around 0.56, and having inflection points close to 0.52 and 0.60.

5 Discussion

This article has showed that regression specifications commonly used in the literature can produce misleading results about the link between match uncertainty and attendance. The regression approach also has two other problems; first, it imposes a specific functional form on the relationship between match balance and attendance; second, its results are hard to interpret, because it is not possible to change match balance without also changing team qualities.

After that, the article has proposed a new approach to examining the link between match uncertainty and attendance that does not rely on regression. Unlike commonly used regression specifications, the proposed method does not find any link between match uncertainty and attendance if the attendance demand is an additively separable function of team qualities (such as in the three simulated data sets) and therefore does not depend on match uncertainty. The non-regression approach also does not assume any specific functional form of the match uncertainty-attendance relationship.

Using data about nine seasons of the English Championship, the proposed method shows that in a pair of matches where both home teams are slight favorites, switching the away teams would decrease the total attendance, while the opposite is true if both home teams are underdogs or strong favorites. However, the impact of such team switches on attendance is just several percent at most. The results are consistent with the uncertainty of outcome hypothesis and suggest that attendance demand is a bell-shaped function of match balance that is maximized if teams of the same quality play against each other (in such matches, home teams are slightly favored due to home advantage). One possible explanation of such a shape could be that there are two groups of potential spectators with different preferences; fans in the first group (seasonal ticket holders, hardcore fans) do not care about match uncertainty and attend all matches if they have free time and no better opportunities, while fans in the second group (occasional spectators) choose to attend only the most interesting matches with one criterion being a proper match balance.

The above results can be directly applied to tournament design; to increase the total attendance of a competition while keeping the number of home and away matches of each team constant, a higher proportion of matches should be played between evenly matched teams. This could be achieved by splitting teams into groups based on team quality instead of on region[8] or by making the tournament design more similar to the Swiss system commonly used in chess. However, the potential attendance increase would likely be small.

---

[8] On the other hand, splitting teams into regional groups would lead to a higher proportion of matches between regional rivals and lower travelling distance between teams. Both of these factors tend to increase attendance (García and Rodríguez, 2009).

Both the small effect size and the bell shape of the attendance demand function further support the claim that the inconsistent results in the previous research could have been caused by misspecified regressions. However, the proposed non-regression approach also has two limitations that result from using historical attendance figures to estimate attendance demand. First, the approach assumes that attendance demand is actually observable and not right-censored (i.e. the stadium is not close to capacity). This is not a problem in the Championship data set, but would be a problem in the Premier League or other top European soccer competitions. A possible solution would be to discard such team combinations where any attendance is close to the corresponding stadium capacity. Second, the non-regression method assumes that other factors influencing match attendance (e.g. day of the week, TV broadcast, distance between teams, weather, or match importance) are not strongly and systematically correlated with match balance. If this assumption does not hold, the results will be biased.

The variable most likely to be correlated with match balance is match importance, i.e. how much a given match result influences the probability of a given season outcome, such as promotion or relegation. If matches between equal-quality teams tend to be more important and higher importance increases attendance, the higher attendance of matches between equal-quality teams would be partly explained by match importance, not by match uncertainty, and the actual effect size would be even lower (the regression approach suffers from the same omitted variable bias problem if match importance is not properly controlled for). Again, a possible solution would be to throw away such team combinations where any match is above some importance threshold.[9] For competition organizers, the distinction between match importance and match uncertainty might not even be relevant; if more matches between equal-quality teams increase attendance, the exact mechanism does not matter.

There are several possible avenues of further research of the link between match uncertainty and attendance. First, the results presented in this article are for one specific competition – English soccer Championship, so the proposed non-regression approach should be applied to soccer competitions in different countries and to different sports to see whether the results stay the same. Second, more attention should be paid to preference-revealing fan behavior during the match. Anecdotally, fans start leaving the stadium prematurely if the score difference is big, especially if the home team is badly losing. The article by Tainsky et al. (2013) is a nice example of this approach applied to TV ratings of NCAA football. Third, fan preferences could be revealed in short series of matches that are essentially one longer match, such as those in the NHL playoffs or European soccer cups; a lower attendance when one team is practically sure to advance to the next round would confirm the uncertainty of outcome hypothesis. Fourth, fans could simply be asked about their preferences related to individual match uncertainty similarly to the stated preferences approach applied to the overall competitive balance by Pawlowski and Budzinski (2013) and Pawlowski (2013).

---

[9] An overview and comparison of methods for calculating match importance can be found in Lahvička (2013). The simplest solution would be to throw away the second half of each season or all team combinations where at least one team was ultimately promoted or relegated.The latter modification applied to the Championship data set does not substantially change the original results.

References

Benz, M.-A., Brandes, L., & Franck, E. (2009, April). Do Soccer Associations Really Spend on a Good Thing? Empirical Evidence on Heterogeneity in the Consumer Response to Match Uncertainty of Outcome. *Contemporary Economic Policy, 27*(2), 216-35.

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

Buraimo, B., & Simmons, R. (2008). Do Sports Fans Really Value Uncertainty of Outcome? Evidence from the English Premier League. *International Journal of Sport Finance*(3), 146-155.

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

Coates, D., & Humphreys, B. R. (2011, June). Game Attendance and Competitive Balance in the National Hockey League. *University of Alberta Department of Economics Working Paper 2011-08*.

Dobson, S., Goddard, J., & Wilson, J. O. (2001). League Structure and Match Attendances in English Rugby League. *International Review of Applied Economics, 15*(3), 335-351.

Forrest, D., & Simmons, R. (2002). Outcome Uncertainty and Attendance Demand in Sport: The Case of English Soccer. *Journal of the Royal Statistical Society - Series D (The Statistician), 51*(2), 229-241.

García, J., & Rodríguez, P. (2009, September). Sports Attendance: A Survey of the Literature 1973-2007. *Rivista di Diritto ed Economia dello Sport*(2), 111-151.

Lahvička, J. (2013). Using Monte Carlo Simulation to Calculate Match Importance: The Case of English Premier League. *Journal of Sports Economics*. doi:10.1177/1527002513490172

Neale, W. C. (1964, February). The Peculiar Economics of Professional Sports: A Contribution to the Theory of the Firm in Sporting Competition and in Market Competition. *The Quarterly Journal of Economics, 78*(1), 1-14.

Pawlowski, T. (2013). Testing the Uncertainty of Outcome Hypothesis in European Professional Football: A Stated Preference Approach. *Journal of Sports Economics*. doi:10.1177/1527002513496011

Pawlowski, T., & Anders, C. (2012). Stadium attendance in German professional football – the (un)importance of uncertainty of outcome reconsidered. *Applied Economics Letters, 19*(16), 1553-1556.

Pawlowski, T., & Budzinski, O. (2013). The Monetary Value of Competitive Balance for Sport Consumers: A Stated Preference Approach to European Professional Football. *International Journal of Sport Finance, 8*(2), 112-123.

Peel, D. A., & Thomas, D. A. (1992). The Demand for Football: Some Evidence on Outcome Uncertainty. *Empirical Economics, 17*(2), 323-331.

Rottenberg, S. (1956, June). The Baseball Players' Labor Market. *Journal of Political Economy, 64*(3), 242-258.

Tainsky, S., Kerwin, S., Xu, J., & Zhou, Y. (2013). Will the real fans please remain seated? Gender and television ratings for pre-game and game broadcasts. *Sport Management Review*. doi:10.1016/j.smr.2013.04.002

Appendix

The left panel of Figure 1 shows the pattern of estimated probabilities of the balanced pair of matches being more attended for various values of *IdealBalance*. This appendix describes how to test whether this pattern of probabilities is statistically significantly different from the pattern of probabilities corresponding to zero impact of match uncertainty on attendance. A simple test statistic is the distance (absolute deviation) of the estimated probability from 0.5 averaged across all *IdealBalance* values. For results depicted in Figure 1, this test statistic equals 0.0569; i.e. the average distance between the line in the graph and the horizontal line at 0.5 is 0.0569.

If the null hypothesis (no impact of match uncertainty on attendance) were true, the balanced pair would be more attended if the sum of error terms (other factors influencing attendance) for the balanced matches were bigger than the sum of error terms for the unbalanced matches. Assuming the normal distribution of error terms, any number of the test statistic values given the null hypothesis can be obtained by keeping all *MatchBalance* values in the data set intact, repeatedly replacing all *Attendance* values by random numbers drawn from a standard normal distribution, and reconstructing the left panel of Figure 1.

In 250 simulations for the Championship data set, the value of the test statistic given the null hypothesis was always lower (mean = 0.0270, standard deviation = 0.0084) than the value of the test statistic using the actual attendance figures (0.0569). This means that such a big deviation from the probability of 0.5 depicted in Figure 1 would be extremely unlikely if there were no effect of match uncertainty on attendance.