Are Fair Weather Fans Affected by Weather? Rainfall, Habit Formation and Live Game Attendance

Qi Ge

Brad R. Humphreys

Kun Zhou

This paper can be found at the College of Business and Economics Working Paper Series homepage:

http://business.wvu.edu/graduate-degrees/phd-economics/working-papers

Are Fair Weather Fans Affected by Weather?

Rainfall, Habit Formation and Live Game Attendance[*]

Qi Ge [†] Brad R. Humphreys [‡] Kun Zhou [§]

August 2017

Abstract: We analyze habit formation in sports attendance utilizing rainfall as an unexpected, transitory shock to attendance costs. Using attendance data from Major League Baseball (MLB) and NOAA weather data, we analyze the impact of variation in game day weather conditions on current and future MLB attendance. The empirical strategy permits identification of both the formation and persistence of habit from exogenous weather shocks. Past adverse weather shocks increase future attendance by about 200 fans per game. Our study contributes to the literature developing empirical evidence of habit formation in the field and provides policy implications for optimal ticket pricing strategies.

Keywords: habit formation, rainfall, attendance
JEL Codes: D03, D12, Z2

1 INTRODUCTION

Researchers have long been interested in demand for live sports attendance and have investigated the role of ticket prices, outcome uncertainty, game importance, local income levels, competitiveness of the home team and opponents, quality of the facility, game day weather conditions, broadcast availability, and other factors. Borland and Macdonald (2003), García and Rodríguez (2009) and Martins and Cró (2016) contain comprehensive reviews of this literature. Behavioral aspects of fan attendance decisions recently attracted attention. For instance, Coates et al. (2014) formulated a game attendance model based on a reference-dependent preferences framework.

Habit formation and persistence is another well-studied area in the behavioral attendance literature. Fans form a habit of attending live sports events if past attendance increases the probability of future attendance. Previous studies (Ahn and Lee, 2007; Lee and Smith, 2008; Pawlowski and Anders, 2012; Martins and Cró, 2016) directly tested whether past (aggregate) attendance leads to future game attendance with mixed results. While empirical attendance models may capture part of fans' habit persistence, much of the research on habit formation and attendance relies on aggregate data at the team- or season-level to identify habituation in attendance, which likely manifests at a more granular game-level.

In addition, previous research also suffers from a potential endogeneity issue that limits researcher's ability to distinguish the formation of an attendance habit from the persistence of habituated fans. Past and current attendance decisions can be correlated with fans' time invariant unobservable personal preferences and traits. Empirically distinguishing habit formation from persistence to provide micro-level evidence of habit formation on attendance requires an exogenous shock to game attendance decisions that is orthogonal to fans' personal traits and past attendance decisions.

This paper utilizes data from Major League Baseball (MLB) and analyzes habit formation in attendance by exploiting game day rainfall as an unexpected, transitory shock to

2

the cost of game attendance. The impact of exogenous adverse weather shocks on fan's attendance decisions represents a source of variation in attendance demand uncorrelated with past attendance and unobserved fan characteristics. This allows us to disentangle habit formation from habit-driven persistence and gauge the magnitude of persistence. In addition, we utilize game-level attendance data, rather than the more aggregated data employed in prior studies in order to provide improved micro-level evidence of habit formation.

We find robust, economically sizable evidence that transitory past weather shocks increase future attendance at MLB games. Specifically, a one standard deviation increase in cumulative rainfall on the previous game day boosts future game attendance by as much as 0.68% or approximately 210 fans on average. The results are robust to the dynamics of attendance decisions and model specifications. The "dual self" model of impulse control developed by Fudenberg and Levine (2006) can explain these results in the case where the long-run attendance habituated self overcomes the short run self's acquisition of the habit of not attending games.

Our paper also contributes to the broader literature developing empirical evidence of habit formation in the field. In addition to a number of laboratory studies (see Duhigg (2012) for a review), recent growth in empirical research on habit formation focuses on consumer decisions made in the field and seeks to address key questions such as whether consumers form habits over time based on past decisions, the persistence of such habits, and under what conditions they will be altered. This paper addresses these questions by analyzing the impact of exogenous weather shocks; our findings regarding live game attendance decisions represent additional evidence in a line of research that includes migrants' preferences toward consumer packaged goods (Bronnenberg et al., 2012), diet variety of migrants (Atkin, 2013), adoption of preventative health behavior (Hussam et al., 2016), voting turnout (Fujiwara et al., 2016), exercise decisions (Humphreys et al. 2015; Royer et al. 2015), charitable donation responses to sports related shocks (Meer, 2013), gasoline demand (Scott, 2012), and home energy consumption (Allcott and Rogers 2014; Ge and Ho 2017).

2 CONCEPTUAL FRAMEWORK

The standard model of rational addiction (Becker and Murphy, 1988) includes time consistent consumers making spending decisions on a good characterized by reinforcement - more consumption of the good in the past increases the marginal utility of consumption today - and tolerance - more consumption in the past decreases the absolute utility from consuming today. In other words, given utility defined over a time path of consumption of an addictive good $c(t)$, the "addictive stock of past consumption" $S(t)$ is increasing in past consumption, and consumption over a non-addictive good $y(t)$, is such that for utility function $U(t) = u[c(t), S(t), y(t)]$, tolerance is defined as $\frac{\partial u}{\partial S} < 0$, and reinforcement is defined as $\frac{\partial c}{\partial S} > 0$. Much of the prior behavioral attendance demand literature follows Becker and Murphy's rational addiction model and predicts that past and future attendance decisions are positively correlated due to the persistence of habit.

A number of economic models featuring habitual attendance at sporting events exist in the literature. Ahn and Lee (2007) develop a two-period model of life cycle attendance at sporting events that includes habitual attendance and investigate the impact of habit formation on price setting. Won and Lee (2008) extend this model to an infinite horizon model. These models show that habit formation may be an important component of attendance decisions, and generate specific predictions about team pricing decisions in the presence of habituated fans. However, these predictions depend on the presence of habit formation among sports fans, which is an empirical question.

In many empirical settings researchers only observe the persistence of habit, presenting an empirical challenge for identifying the formation of habit. Measures of persistence of habit in sports attendance have been limited to the inclusion of lagged attendance in most previous research, a practice that has well-known problems (Dawson et al., 2000). We avoid this problem by using weather shocks to identify habit formation. Weather shocks introduce cues that may exogenously trigger the formation of (a possibly different) habit. For example,

4

Fujiwara et al. (2016) document that rainfall not only reduces current voter turnout but also negatively affects future voter turnout.

Intuitively, game day precipitation increases the cost of attending a game through logistical inconveniences (travel takes longer because rain slows traffic and walking from the parking lot to the stadium in the rain is unpleasant) and a less pleasant game viewing experience. This is especially true for open air stadiums.[1] Price and Sen (2003) empirically document a negative impact of bad weather on *current* attendance. We posit that adverse weather on a given game day will not only dampen current game attendance but also possibly activate the habit of not attending games, i.e., a current weather shock may hinder fans (at least those fair weather ones) from attending *future* games.

On the other hand, the existing persistence of game attendance may interfere with the formation of the habit of not attending games. The stronger the persistence of the habit of attending games, the more likely it is that a fan who planned to attend a game on Tuesday but did not attend that game because of a negative weather shock would instead attend the game on Wednesday. The possible presence of strong persistence in the attendance habit clouds the overall empirical prediction about the impact of past weather shocks on current game attendance. These decisions can be motivated by the "dual self" model of impulse control developed by Fudenberg and Levine (2006).

Finally, studies such as Nelson and Meyvis (2008) document how interruptions during an hedonic experience help restore the intensity of the experience and increase enjoyment. We argue, in the same vein, that a transitory weather shock may temporarily interrupt the habit of attending games, which may in turn result in an urge to attend future games. As the stock of such urges accumulate, a fan may become more likely to make an active decision to satiate the urge by attending the next game. In other words, it is possible that past weather shocks may in fact lead to increases in current attendance.

---

[1]23 out of the 30 MLB teams play in open air home stadiums. MLB games are typically not affected by light to moderate rain. In case of a postponed or canceled game due to heavy rain or other adverse weather conditions, teams have "rainout" policies that allow fans to exchange tickets to the canceled game for a future game. We focus only on completed games.

3 Data and Empirical Strategy

3.1 Data

The data come from several sources. Our primary source is the MLB attendance data used by Coates et al. (2014). We utilize game day attendance data for MLB games and update their data to cover approximately 29,000 regular season games from the 2005 through 2016 MLB regular season. The official game attendance defined by MLB is the number of tickets sold for a particular game, rather than the actual number of people who attended the game. The actual turnout rate is not publicly available. Teams sell season tickets (packages of tickets to specific sets of games including all 81 home games), advance sale single game tickets, and single game tickets on game day. A recent case study from MLB reports that about 30% of tickets were sold on game day (Xu et al., 2016). Following the practice in the literature, we assume that "attendance" reflects the number of people attending each game.

In addition, we incorporate MLB moneyline betting odds data from `www.covers.com` to control for pre-game expectations of game outcomes. We convert the money line odds to the probability that the home team will win each game using the standard formula.

The game day weather data come from the National Oceanic and Atmospheric Administration (NOAA) and contain hourly accounts of outdoor temperature and precipitation from the weather station closet to each MLB stadium in the sample matched by GPS coordinates. Data on hourly local precipitation intensity, in inches per hour, are collected for each game day in our dataset. We mainly focus on rainfall as the weather shock of interest given that the MLB regular season typically runs from April through September and it is less common to observe other adverse weather conditions such as snow, ice or sleet during this period. Besides historical local weather data, we also collect the one-day forecast of average temperature and total precipitation amount from each weather station for each game day.[2]

---

[2]The forecast data are from the NOAA Global Ensemble Reforecast Data Set, available at `https://www.esrl.noaa.gov/psd/forecasts/reforecast2/`. Note that NOAA's Reforecast data do not contain

Each game in the dataset can be uniquely identified by the home team, the visiting team, and the date of the game. Due to the presence of doubleheaders, where the same pair of teams play two games on the the same day, knowing the date and the teams involved may not uniquely identify a game. Given that our identification strategy is based on game day weather conditions, we restrict our sample to those game days when a single game was played and, in the case of a doubleheader, the first game of the doubleheader. Previous rainfall is the amount of rainfall recorded on the day of the previous home game played.

Table 1 provides summary statistics for the main variables of interest. Average attendance at MLB games in the sample is approximately 31,000 fans per game with a significant heterogeneity across time and cities. We also observe a sizable variation in the game day average outdoor temperatures and recent performances of both the home and visiting teams. The game day average cumulative rainfall is approximately 0.07 inch and the probability of raining on a given game day is about 8% with understandably large geographical and time variations. On the other hand, heavy rain ($\geq$0.05 inch/hour) is relatively rare on game days, possibly because of the team's rainout policy. The MLB follows a compact game schedule with an average of approximately two days between any two home games. The high frequency of games combined with large variations in rainfall will facilitate our identification strategy. Weather forecast data on average match well with the actual historical weather data with respective correlations of approximately 0.50 and 0.88 for precipitation and outdoor temperature.

3.2 EMPIRICAL STRATEGY

Our empirical strategy resembles the approach used by Fujiwara et al. (2016). We exploit local variation in game day weather conditions and track the attendance at home games over time, controlling for a set of well-studied determinants of attendance demand used

the actual historical published forecasts. Rather, they are based on re-generation of forecasts utilizing the same forecast model used in the past. NOAA states that the Reforecast dataset is statistically consistent with the current weather forecast system. We believe that the Reforecast dataset can serve as a good proxy for the actual weather forecast known by fans when deciding whether to attend a game.

Table 1: Summary Statistics

Variable	Obs.	Mean	Std. Dev.	Min	Max
Attendance	28,871	30,912	10,450	6017	57,405
Day game	28,871	0.32	0.47	0	1
$\overline{Runs}_H$	27,970	4.47	1.45	0.2	12.2
$\overline{Runs}_V$	27,970	4.45	1.47	0.4	13
Home win probability	28,109	0.55	0.08	0.25	0.81
Days between games	28,511	2.21	2.99	1	16
Daily cum. precipitation (inches)	28,628	0.07	0.20	0	4.47
Rain indicator	28,628	0.08	0.27	0	1
Heavy rain indicator	28,628	0.01	0.12	0	1
Average outdoor temperature (°F)	28,871	71.59	12.92	12.44	107.73
Forecasted daily cum. precipitation (inches)	28,835	0.10	0.22	0	3.76
Forecasted average outdoor temperature (°F)	28,850	67.01	11.51	21.84	103.41

$\overline{Runs}$ is the average runs scored per game over the last 5 games for the home (H) and visiting (V) team. *Home win probability* is the home team's pre-game predicted probability of winning based on final Las Vegas money line betting odds. Weather forecast data are based on one-day forecasts.

in previous empirical research. We first consider the following empirical specification that captures the contemporaneous impact of rainfall on attendance:

$$\ln(attendance)_{ct} = \beta \cdot rain_{ct} + X'_{ct}\xi + \tau_t + \eta_c + \epsilon_{ct} \tag{1}$$

where $\ln(attendance)_{ct}$ is the natural logarithm of attendance at MLB game played on date t in team c's home stadium, $rain_{ct}$ is the cumulative rainfall on game day t in the vicinity of team c's home stadium,[3] τ_t is a time fixed effect capturing factors that affect all MLB games played on date t, η_c is a team fixed effect for team c, and X_{ct} is the set of control variables that affect game-level attendance based on factors identified in the existing literature on attendance demand, including pre-game expectations of game outcomes and recent performance of both the home and visiting teams.

In addition, we control for the number of days since the last home game and the average

---

[3]We also consider alternative rainfall measures such as hourly precipitation close to game start as well as forecasted rainfall in Section 4.2.

game day outdoor temperature.[4] To capture possible nonlinearity in the relationship between temperature and attendance, we also include a quadratic term for the average game day outdoor temperature - day games with extreme temperatures may not offer a pleasant experience to fans. Finally, estimated standard errors are cluster-corrected at home-team-city level.

To capture the impact of past rainfall on current attendance, we augment Equation (1) by including a lagged rainfall term:

$$\ln(attendance)_{ct} = \beta \cdot rain_{c,t} + \delta \cdot rain_{c,t-1} + X'_{c,t}\xi + \tau_t + \eta_c + \epsilon_{ct} \tag{2}$$

This model permits estimation of the average effect of rainfall on attendance along with the effects of past rainfall on attendance, captured by the lagged rainfall variable. The lagged rainfall variable reflects the cumulative amount of rain recorded on the last day a home game was played in the stadium. Rainfall is plausibly exogenous to unobservable fan characteristics and unobservable factors affecting demand for attendance at specific MLB games. This specification thus represents a reduced form model of the impact of rain on attendance. As discussed in Section 2, the prediction of the sign of δ, the parameter on the lagged rainfall variable, cannot be signed.

It is possible that the impact of past rainfall on future attendance decisions may extend beyond the previous game. We investigate the possibility in a separate model by introducing additional lags of the rainfall variable, e.g., the cumulative rainfall from two and three home games ago. Previous studies on habit formation in live game attendance often rely on directly regressing current attendance on past attendance. This empirical strategy may suffer from an endogeneity problem. Since game day rainfall is arguably correlated with the game day attendance decision but exogenous to personal traits as well as past attendance decisions, an alternative empirical framework consistent with prior research would involve using past

---

[4]Games may be shortened because of rain. MLB rules state that an official game lasts 4.5 innings if the home team is leading the game, and 5 innings if the visiting team is leading (i.e., the home team must complete its at-bats in the 5th inning). Games played in rainy conditions could be shorter than 9 innings.

rainfall shocks as instruments for past attendance.[5] Equation 2 is our preferred specification because it represents a reduced form specification of habit formation that reflects the impact of past weather shocks and includes the dynamics of past attendance decisions.

We control for dynamics in attendance decisions by including the natural log of attendance at the previous home game as an independent variable. Finally, to account for possible spurious correlation, we also perform a falsification test similar to Miguel et al. (2004), where future rainfall is included as an independent variable in the regression model. We posit that the estimated coefficient on the future rainfall variable should be zero in a correctly specified model.

4   RESULTS AND DISCUSSION

4.1   MAIN FINDINGS

Table 2 contains our main empirical findings. Column (1) presents the results for the basic regression model specification that only includes game day and lagged precipitation and team fixed effects. Contemporaneous (game day) rainfall reduces attendance while past rainfall increases current attendance. Column (2) introduces time fixed effects (year, month and day of week); lagged rainfall still has a positive impact on current attendance while the impact of contemporaneous weather shocks falls. Column (3) includes a set of factors that may influence attendance decisions as documented in prior attendance studies, including past performance of the home and visiting teams, pre-game expectations of game outcome and game day temperature (and its quadratic). We also control for the number of innings in each game as well as the number of days since last home game.[6] The sign and magnitude of

---

[5]As a robustness check, we estimate this panel IV specification; the results are discussed in Footnote 7.

[6]The attendance demand literature also typically controls for the total cost of attendance. Total cost of attendance data can be found in the Fan Cost Index (FCI), collected and published in *Team Marking Report*. Our main models exclude the FCI since the FCI data are at the season-level and our identification is at the game-level. However, when we add the natural log of FCI as a control variable to the existing model specifications, we obtain identical results to those in Table 2.

the estimated parameter on the lagged rainfall variables are similar to those in Columns (1) and (2). The parameter estimate on contemporaneous rainfall becomes insignificant.

The results indicate that a one standard deviation increase in lagged cumulative game day rainfall will lead to a 0.42% increase in attendance or an average of approximately 130 additional fans, which we argue is economically significant. In terms of the control variables, we find that recent home team performance, in terms of runs scored, and home team pre-game predicted probability of winning both help boost attendance, which is expected and in line with previous results. In addition, game day outdoor temperature is also positively correlated with attendance. Although there seems to be evidence for a nonlinear impact of outdoor temperature, its magnitude is small across all model specifications.

We control for the dynamics of attendance decisions in Column (4) by including lagged attendance as an explanatory variable; the sign and significance of current and lagged rainfall remains the same with larger magnitudes in this model. In particular, the coefficient on the lagged rainfall variable is 0.034, implying that a one standard deviation increase in the lagged cumulative game day rainfall increases current attendance by as much as 0.68% or approximately 210 fans. Finally, in Column (5), we introduce a future rainfall variable that presumably should not affect current attendance decisions if our model is correctly specified. The estimated coefficient on the future rainfall is more than three times smaller than the current or lagged rainfall coefficient and is not statistically different from zero.

We further investigate the extent to which past rainfall shocks affect future attendance decisions using the same model specifications as in Table 2 and introducing two additional lags of game day rainfall. That is, we consider the additional impact of weather shocks from two and three home games ago. Table 3 shows parameter estimates on current and lagged precipitation variables. Rainfall on the three most recent game days all have positive and significant impacts on current game attendance, though once the dynamics of game attendance is taken into account in Column (4), only precipitation on the two most recent game days affects current attendance.

Table 2: Impact of Weather Shocks - Main Model Specifications

	(1)	(2)	(3)	(4)	(5)
Current rainfall	-0.022**	-0.014*	-0.004	-0.024***	-0.024***
	(0.009)	(0.008)	(0.007)	(0.006)	(0.009)
Lagged rainfall	0.022**	0.018**	0.021**	0.034***	0.038***
	(0.008)	(0.008)	(0.008)	(0.007)	(0.007)
Next day rainfall					0.008
					(0.009)
$\overline{Runs}_H$			0.006***	0.004***	0.004***
			(0.002)	(0.001)	(0.001)
$\overline{Runs}_V$			0.002	-0.000	0.001
			(0.002)	(0.001)	(0.001)
Home win probability			0.385***	0.111***	0.086**
			(0.087)	(0.036)	(0.038)
Avg. temperature			0.013***	0.009***	0.008***
			(0.003)	(0.002)	(0.002)
Avg. temperature^2			-0.001**	-0.001***	-0.001***
			(0.000)	(0.000)	(0.000)
Days b/w games			0.003***	-0.000	0.000
			(0.000)	(0.001)	(0.001)
Lagged attendance				0.591***	0.557***
				(0.022)	(0.025)
Observations	28.318	28.318	26.763	26.612	20.968
Team FE	Yes	Yes	Yes	Yes	Yes
Time FE	No	Yes	Yes	Yes	Yes

The dependent variable across all specifications is $log(attendance)$. *Current rainfall* is cumulative rainfall on game day. *Lagged rainfall* is cumulative rainfall on the previous game day. $\overline{Runs}$ is the average runs scored per game over the last 5 games for the home (H) and visiting (V) team. *Home win probability* is the home team's pre-game predicted probability of winning based on final Las Vegas money line betting odds. *Average temperature* is the game day average outdoor temperature. Time fixed effects include year, month and day of week. Robust standard errors in parentheses are clustered at the city level. Very small point estimates and standard errors were rounded to 0.000. *** p<0.01, ** p<0.05, * p<0.1.

From Column (5), this specification continues to pass the falsification test. Overall, the results in Table 3 lend strong support to the conclusion that that the impact of past rainfall extends beyond just the most recent game day.[7]

---

[7]We also estimate a panel IV specification using past lags of rainfall shocks as instruments for past attendance. The IV estimates suggest that, consistent with persistent habit, past attendance is positively correlated with current attendance with an estimated coefficient of 1.016 (significant at the 1% level). The first stage F-statistic is 17.77, which is within 5% of the maximal relative bias according to the Stock-Yogo weak instrument test. The instruments also pass the overidentifying restrictions test (p-value: 0.378).

Table 3: Impact of Weather Shocks - Multiple Weather Lags

	(1)	(2)	(3)	(4)	(5)
Current rainfall	-0.024**	-0.016**	-0.005	-0.024***	-0.024***
	(0.009)	(0.008)	(0.007)	(0.007)	(0.009)
Past rainfall - 1st lag	0.016**	0.012*	0.016**	0.031***	0.036***
	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)
Past rainfall - 2nd lag	0.034***	0.032***	0.028***	0.019***	0.019**
	(0.009)	(0.009)	(0.007)	(0.006)	(0.007)
Past rainfall - 3rd lag	0.034***	0.031***	0.017**	0.003	0.003
	(0.011)	(0.009)	(0.007)	(0.005)	(0.004)
Next day rainfall					0.007
					(0.009)
Lagged attendance				0.591***	0.557***
				(0.022)	(0.025)
Observations	28.254	28.254	26.762	26.611	20.968
Team FE	Yes	Yes	Yes	Yes	Yes
Time FE	No	Yes	Yes	Yes	Yes
Controls	No	No	Yes	Yes	Yes

The dependent variable across all specifications is the natural log of attendance. *Current rainfall* is cumulative rainfall on game day. *Past rainfall - ith lag* is cumulative rainfall on the *i*th previous game day. Control variables for Columns (3) to (5) are the same as those in Table 2. Time fixed effects include year, month and day of week. Robust standard errors in parentheses are clustered at the city level. *** p<0.01, ** p<0.05, * p<0.1.

4.2 Robustness Checks[8]

We perform a series of robustness checks on our main model specifications. First, in line with prior literature on live game attendance, we reestimate our main model specifications using a panel tobit model to take into account right censored games (sold out games). These estimates yield similar results to those on Tables 2 and 3.[9] Previous research (Meehan Jr et al., 2007) shows that baseball games are not usually subject to strong right-censoring, and in our data the mean and median attendance rates are approximately 70% of capacity with only 11.56% of the games being completely sold out.

Second, we estimate the models separately for home stadiums with and without domes since the live game viewing experience in open air stadiums is presumably more affected by weather shocks than in domed stadiums. These estimates are shown on Table 4. Overall,

---

[8]Results presented in this section are based on empirical specifications derived from Equation (2) with only one lag of the game day rainfall variable. We also perform the same set of the robustness checks with multiple lags of game day rainfall, and the results are identical. These results are available upon request.

[9]The panel tobit results are available upon request.

the results for open air stadiums are qualitatively identical to those in Table 2, with lagged rainfall displaying a larger magnitude when dynamics of attendance decision is taken into account in Columns (4) and (5). We also conduct the same analysis for home stadiums with domes and do not find statistically significant coefficient estimates on the lagged precipitation variable as well as most of the control variables of interest across all specifications.

In fact, there is weak evidence that rainfall may boost contemporaneous attendance as exhibited in Columns (3) and (5). This result could reflect the idea that attending a baseball game indoors on a rainy day is a substitute for other outdoor recreational activities that become unatractive on rainy days, like going to a park or amusement park or going for a walk. These results confirm that while rainy weather may inconvenience fans traveling to and from the stadium, the live game viewing experience seems to be a more important channel through which weather shocks affect attendance decisions.

Finally, we consider different measures of rainfall in our empirical models, which is broadly related to the temporal dimension of fans' game attendance decisions. Some fans, especially casual and fair-weather ones, may make game attendance decisions conditional on weather conditions just a few hours prior to the game. Alternatively, their attendance decisions may be based on weather forecasts. To address this concern, we follow the same empirical model specifications as in Table 2 but consider different proxies for game day precipitation. Specifically, we consider the cumulative rainfall two hours before the game, four hours before the game, and the forecasted daily cumulative rainfall amount from the day before the game. The results are presented in Table 5.[10] There is no evidence that fans' attendance decisions are based on weather conditions just hours before the game. Rather, they seem to pay attention to weather forecasts and make attendance decisions accordingly. In fact, the magnitudes of the impact of forecasted rainfall in Columns (5) and (6) are decisively larger than the comparable estimates of the contemporaneous rainfall variable in Table 2. This suggests that even casual fans may be forward-looking, which is consistent with the basic

---

[10]Note that in Table 5 the lagged rainfall variable is still based on the actual precipitation amount from the previous game day.

Table 4: Robustness Check - Open Air vs. Domed Stadiums

	(1)	(2)	(3)	(4)	(5)
OPEN AIR STADIUMS					
Current rainfall	-0.028**	-0.022**	-0.013*	-0.033***	-0.035***
	(0.011)	(0.009)	(0.007)	(0.007)	(0.008)
Lagged rainfall	0.021**	0.018*	0.021**	0.039***	0.043***
	(0.009)	(0.009)	(0.009)	(0.009)	(0.009)
Next day rainfall					0.009
					(0.010)
Lagged attendance				0.596***	0.566***
				(0.028)	(0.032)
Observations	22.742	22.742	21.514	21.381	16.789
DOMED STADIUMS					
Current rainfall	0.002	0.014	0.029**	0.015	0.020*
	(0.011)	(0.010)	(0.009)	(0.010)	(0.009)
Lagged rainfall	0.025	0.018	0.015	0.008	0.017
	(0.019)	(0.016)	(0.012)	(0.015)	(0.018)
Next day rainfall					0.003
					(0.011)
Lagged attendance				0.548***	0.501***
				(0.035)	(0.040)
Observations	5.576	5.576	5.249	5.231	4.179
Team FE	Yes	Yes	Yes	Yes	Yes
Time FE	No	Yes	Yes	Yes	Yes
Controls	No	No	Yes	Yes	Yes

The dependent variable across all specifications is the natural log of attendance. *Current rainfall* is cumulative rainfall on game day. *Lagged rainfall* is cumulative rainfall on the previous game day. Control variables for Columns (3) to (5) are the same as those in Table 2. Time fixed effects include year, month and day of week. Robust standard errors in parentheses are clustered at the city level. *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

setup in a Becker-Murphy style rational addiction model.

4.3 DISCUSSION

Our findings provide evidence that negative weather shocks affect future attendance decisions, likely through their impact on habit formation. Intriguingly, while contemporaneous rainfall reduces current attendance, past rainfall shocks increase future attendance, and the effect is robust to different model specifications and controls. In other words, rainfall and the resulting perceived unpleasant game viewing experience may introduce a temporary and non-lasting break in the persistence of attendance. This is in contrast to the findings in Fu-

Table 5: Robustness Check - Alternate Rainfall Measures

	Two-hour rainfall		Four-hour rainfall		Forecasted rainfall	
	(1)	(2)	(3)	(4)	(5)	(6)
Current rainfall	-0.050	-0.041	-0.005	-0.019	-0.034***	-0.030***
	(0.040)	(0.026)	(0.017)	(0.015)	(0.010)	(0.007)
Lagged rainfall	0.021**	0.031***	0.020**	0.031***	0.026***	0.036***
	(0.008)	(0.007)	(0.008)	(0.007)	(0.008)	(0.007)
$\overline{Runs}_H$	0.006***	0.004***	0.006***	0.004***	0.006***	0.004***
	(0.002)	(0.001)	(0.002)	(0.001)	(0.002)	(0.001)
$\overline{Runs}_V$	0.002	-0.000	0.002	-0.000	0.002	-0.000
	(0.002)	(0.001)	(0.002)	(0.001)	(0.002)	(0.001)
Home win probability	0.385***	0.111***	0.385***	0.111***	0.387***	0.112***
	(0.087)	(0.036)	(0.087)	(0.036)	(0.087)	(0.036)
Avg. temperature	0.013***	0.009***	0.013***	0.008***	0.013***	0.008***
	(0.003)	(0.002)	(0.003)	(0.002)	(0.003)	(0.002)
Avg. temperature^2	0.000**	0.000**	0.000**	0.000***	0.000***	0.000***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Days b/w games	0.003***	-0.000	0.003***	-0.000	0.003***	-0.000
	(0.000)	(0.001)	(0.000)	(0.001)	(0.000)	(0.001)
Lagged attendance		0.591***		0.591***		0.591***
		(0.022)		(0.022)		(0.022)
Observations	26.760	26.609	26.760	26.609	26.734	26.583
Team FE	Yes	Yes	Yes	Yes	Yes	Yes
Time FE	Yes	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes

The dependent variable across all specifications is the natural log of attendance. *Current rainfall* is proxied by the cumulative rainfall two hours before the game, four hours before the game and forecasted the day before, respectively. *Lagged rainfall* is the actual cumulative rainfall on the previous game day. Control variables for Columns (3) to (5) are the same as those in Table 2. Time fixed effects include year, month and day of week. Robust standard errors in parentheses are clustered at the city level. *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

jiwara et al. (2016), where voters form a lasting habit of not voting as a result of transitory weather shocks.

Our results reflect two possible underlying mechanisms. First, the empirical evidence may imply that, while some fans start forming a new habit of not attending games due to a weather shock, the persistence of the habit of attending games dominates. In this case, the evidence suggests that habit persistence for attending sports events is very strong; strong enough to overcome a negative temporary shock to the cost of attending a game.

Alternatively, consistent with the notion that interruptions during a hedonic experience may help restore the intensity of the experience and increase its enjoyment (Nelson and Meyvis, 2008), it is possible that the weather shock activates a stock of urge to attend future games, which leads to increased future attendance. Based on the results from model

16

specifications that control for the persistence of attendance, we believe that this mechanism could be more likely. Results from Table 3 shed further light toward the urge explanation as negative past weather shocks beyond the previous game day continue to pose additional positive impact on future game attendance.

5   Conclusions

In this paper, we utilize exogenous weather shocks to identify fans' habit formation separately from its persistence. We find evidence that rainfall temporarily introduces a break in attendance persistence. However, fans do not necessarily form a habit of not attending games in response to this shock. Instead, we find that past rainfall in fact promotes future game attendance. The habit of attending MLB games appears to be highly persistent at the game level.

Our results indicate that habit-based persistence exists in attendance at MLB games, and that the strength of this persistence can overcome negative weather shocks that increase the current cost of attendance. This represents stronger evidence of habit persistence than developed in previous research using only lagged attendance to capture the effects of habit-based persistence and aggregate team- and season level data.

If correctly identified and economically significant, the presence of habit formation in attendance has important economic and policy implications for optimal ticket pricing strategies. For example, past research (Lee, 2006) suggests that ticket prices are lower than those predicted by the standard neoclassical profit-maximizing model. One possibility is that the total financial cost of attending games is more than just the ticket price due to the cost of transportation, lodging, and meals, etc (Krautmann and Berri, 2007). On the other hand, our results imply that the high cost of attending a particular game may not always deter fans from attending future games due to the persistence of habit. It is thus important to take into account behavioral factors such as fans' attendance habits when constructing optimal

ticket pricing strategies.

The results also have important implications for dynamic ticket pricing, a recent innovation in MLB ticket pricing where teams charge different prices for different games based on current conditions, including recent team success, opponent quality, and current weather conditions. Evidence suggests that dynamic pricing can increase ticket revenues, but current dynamic pricing models do not account for habit formation in attendance, or past weather shocks affecting current attendance (Xu et al., 2016). Including persistent attendance habituation and past weather conditions to dynamic pricing models can further enhance team revenues from ticket sales.

References

Ahn, S., and Lee, Y. H. (2007). "Life-cycle demand for Major League Baseball." *International Journal of Sport Finance*, *2*(2), 79–93.

Allcott, H., and Rogers, T. (2014). "The short-run and long-run effects of behavioral interventions: Experimental evidence from energy conservation." *American Economic Review*, *104*(10), 3003–3037.

Atkin, D. (2013). "Trade, tastes, and nutrition in India." *American Economic Review*, *103*(5), 1629–63.

Becker, G. S., and Murphy, K. M. (1988). "A theory of rational addiction." *Journal of Political Economy*, *96*(4), pp. 675–700.

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

Bronnenberg, B. J., Dubé, J.-P. H., and Gentzkow, M. (2012). "The evolution of brand preferences: Evidence from consumer migration." *American Economic Review*, *102*(6), 2472–2508.

Coates, D., Humphreys, B. R., and Zhou, L. (2014). "Reference-dependent preferences, loss aversion, and live game attendance." *Economic Inquiry*, *52*(3), 959–973.

Dawson, A., Downward, P., et al. (2000). "Measuring habit persistence effects in attendance at professional team sports encounters: A cautionary note." *Economic Issues Journal Articles*, *5*(1), 37–40.

Duhigg, C. (2012). *The Power of Habit: Why We Do What We Do in Life and Business.* Random House.

Fudenberg, D., and Levine, D. K. (2006). "A dual-self model of impulse control." *American Economic Review*, *96*(5), 1449–1476.

Fujiwara, T., Meng, K., and Vogl, T. (2016). "Habit formation in voting: Evidence from rainy elections." *American Economic Journal: Applied Economics*, (4), 160.

García, J., and Rodríguez, P. (2009). "Sports attendance: A survey of the literature 1973-2007." *Rivista di Diritto e di Economia dello Sport*, *5*(2), 112–151.

Ge, Q., and Ho, B. (2017). "Energy consumption and habit formation: Evidence from high frequency thermostat usage data." Working paper.

Humphreys, B., Ruseski, J., and Zhou, L. (2015). "Physical activity, present bias, and habit formation: Theory and evidence from longitudinal data." Working Papers 2015-6, University of Alberta, Department of Economics.

Hussam, R., Rigol, N., Reggiani, G., and Rabbani, A. (2016). "Habit formation and rational addiction: A field experiment in handwashing." Working paper.

Krautmann, A. C., and Berri, D. J. (2007). "Can we find it at the concessions? Understanding price elasticity in professional sports." *Journal of Sports Economics*, *8*(2), 183–191.

Lee, Y. H. (2006). "The decline of attendance in the Korean Baseball League: The Major League effects." *Journal of Sports Economics*, *7*(2), 187–200.

Lee, Y. H., and Smith, T. G. (2008). "Why are Americans addicted to baseball? An empirical analysis of fandom in Korea and the United States." *Contemporary Economic Policy*, *26*(1), 32 – 48.

Martins, A. M., and Cró, S. (2016). "The demand for football in Portugal: New insights on outcome uncertainty." *Journal of Sports Economics*, Forthcoming.

Meehan Jr, J. W., Nelson, R. A., and Richardson, T. V. (2007). "Competitive balance and game attendance in Major League Baseball." *Journal of Sports Economics*, *8*(6), 563–580.

Meer, J. (2013). "The habit of giving." *Economic Inquiry*, *51*(4), 2002–2017.

Miguel, E., Satyanath, S., and Sergenti, E. (2004). "Economic shocks and civil conflict: An instrumental variables approach." *Journal of Political Economy*, *112*(4), 725–753.

Nelson, L. D., and Meyvis, T. (2008). "Interrupted consumption: Disrupting adaptation to hedonic experiences." *Journal of Marketing Research*, *45*(6), 654–664.

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

Price, D. I., and Sen, K. C. (2003). "The demand for game day attendance in college football: An analysis of the 1997 Division 1-A Season." *Managerial and Decision Economics*, (1), 35.

Royer, H., Stehr, M., and Sydnor, J. (2015). "Incentives, commitments, and habit formation in exercise: Evidence from a field experiment with workers at a Fortune-500 company." *American Economic Journal: Applied Economics*, *7*(3), 51–84.

Scott, K. R. (2012). "Rational habits in gasoline demand." *Energy Economics*, *34*(5), 1713–1723.

Won, D. C., and Lee, Y. H. (2008). "Optimal dynamic pricing for sports games with habitual attendance." *Managerial and Decision Economics*, *29*(8), 639–655.

Xu, J. J., Fader, P., and Veeraraghavan, S. K. (2016). "The revenue impact of dynamic pricing policies in Major League Baseball ticket sales." Working paper, SSRN, https://ssrn.com/abstract=2796407.