Journal of Quantitative Analysis in Sports

Volume 6, *Issue* 1 | 2010 | *Article* 4

Do Fans Matter? The Effect of Attendance on the Outcomes of Major League Baseball Games

Erin E. Smith, *New York University*
Jon D. Groetzinger, *SeatGeek*

Recommended Citation:
Smith, Erin E. and Groetzinger, Jon D. (2010) "Do Fans Matter? The Effect of Attendance on the Outcomes of Major League Baseball Games," *Journal of Quantitative Analysis in Sports*: Vol. 6: Iss. 1, Article 4.

DOI: 10.2202/1559-0410.1192

Do Fans Matter? The Effect of Attendance on the Outcomes of Major League Baseball Games

Erin E. Smith and Jon D. Groetzinger

Abstract

We examine the role of attendance in home-field advantage for Major League Baseball, using a dataset of all MLB games played from 1996 to 2005. Using two-stage least squares, we find that attendance has a significant effect on the home-field advantage. Our results indicate that a one standard deviation increase in attendance results in a 4% increase in the likelihood of a home team win. We also find that if attendance as a percent of stadium capacity were to increase by 48%, we would expect the home team's run differential to increase by one run. We show that the additional home-field advantage is driven by increased home team performance.

KEYWORDS: baseball, attendance, fans, home-field advantage

Author Notes: We wish to thank John Milliken for his contribution to the study, Douglas Staiger and William Greene for their helpful comments and suggestions, and Dan Rascher for providing a dataset used in the study. We would also like to thank the editor and anonymous referee whose comments have greatly improved the paper.

Erratum

Please note that the following statement has been retracted from Page 4:

``Bradbury (2007) shows racial discrimination is less likely to occur when the umpires are monitored by an electronic pitch tracking system called QuesTec."

There is no argument in the paper by Bradbury that such discrimination is likely to occur.

Introduction

It is well-established that a home-field advantage exists in professional sports. Carmichael and Thomas (2005) report that the home team win percentage is 53.5% for baseball, 57.3% for football, 61.1% for ice hockey, 64.4% for basketball, and 69% for soccer. There are multiple sources of this advantage; Courneya and Carron (1992) suggest that the toll of traveling and the psychological effect of playing in an unfamiliar setting may hurt the visiting team. They also discuss, along with Pace and Carron (1992), an additional "comfort factor" benefit for the home team—staying in a hotel without a home-cooked meal may have a negative psychological impact. Schwartz and Barsky (1977) point out that baseball affords two additional advantages—the home team bats after the visiting team (allowing more strategic options) and baseball field dimensions are not standardized, so the home team is more familiar with the layout.

In addition to the edge provided by the above aspects of home-field advantage, there is a prevailing belief that the enthusiasm of the fans helps the home team win. Post-game interviews often feature athletes crediting the fans for giving them a psychological edge. Writers and commentators sometimes suggest that a team performed better than it should have because of extraordinary fan support. During a 2006 end-of-season run, Penn State's Basketball Coach, Rene Portland, beseeched fans, "I really hope that everybody…brings a friend because we need the crowds to make this team go strong down these last seven games" (Blau, 2006). The San Diego Chargers have deliberately changed ticket policies and prices to make it more difficult for fans in Oakland (home of the Raiders, the Chargers' biggest rival) to buy tickets. According to Chargers spokesman, Bill Johnson, "More Chargers fans are going [to] help us win. It creates a stronger home-field advantage for us" (Acee, 2002).

Despite common perception, the hypothesis that fans affect the outcome of games has undergone little formal testing, and none specific to baseball. In this paper, we examine the effect of fans on home and away team performance in Major League Baseball (MLB). We use attendance as a measure of fan support and show that a one standard deviation increase in attendance leads to an additional half a run and a 4% increase in the likelihood of a home win. We then test for several alternative explanations of this effect. We hypothesize that the additional attendance based home-field advantage could be due to (1) increased disparity in the relative performance of the home and away teams, (2) a change in managerial decisions, or (3) the possibility that umpires will become biased (or more biased) in favor of the home team when attendance is high. We find supporting evidence for the first two explanations but find no evidence of attendance-generated umpire bias.

To further motivate our study, we consider the impact of home-field advantage on optimal ticket pricing. Silver (2006) finds that the revenue generated from one additional win in a season is approximately $1.8 million for MLB teams. While a literature on ticket pricing exists, that literature does not test whether higher attendance could boost revenue via the effect of attendance on winning percentage. Our results suggest that it could be profitable for franchises to decrease ticket prices in order to generate more wins.

The contribution of our paper is three-fold. First, we quantify the home-field advantage generated by fan support. Second, we show how attendance impacts relevant performance statistics. Third, we determine that this advantage is driven primarily by superior performance of the home team and not from fan generated umpire bias. Forth, we show that even using conservative estimates of demand elasticity, it could be profitable for managers to decrease ticket prices in order to increase the likelihood of a home win.

Prior Research

There is overwhelming evidence that a statistically significant home-field advantage pervades nearly all professional sports. Courneya and Carron (1992) provide the most comprehensive survey of research on home-field advantage. They use binomial tests to find that the chance of winning at home is significantly different from 50% in all major sports. The effect is greatest for indoor sports like ice hockey and basketball and somewhat smaller for outdoor sports like baseball and football.

Many researchers have sought to determine why home-field advantage exists. Nevill and Holder (1999) describe home-field advantage as deriving from four causes: rules, travel factors, familiarity with the home playing field, and fan support. Generally, researchers have had limited success attributing a quantifiable home-field advantage to any of these factors.

Researchers find that travel factors such as the number of time zones crossed by the visiting team or the number of games the team has played on the road may contribute to home-field advantage, but the effect is small. Pace and Carron (1992) find that travel-related factors explain only 1.5% of the home-field advantage for ice hockey. Courneya and Carron (1991) find a similar result for basketball. They conclude that travel accounts for only 1.2% of the home-field advantage.

There is also only weak evidence that familiarity with the stadium and playing surface gives an advantage to the home team. Barnett and Hilditch (1993) find that English soccer teams who played on artificial turf enjoyed an abnormally

DOI: 10.2202/1559-0410.1192

large home-field advantage in the 1980s*. Other researchers, however, find that the home-field playing surface does not contribute to the home-field advantage. Clarke and Norman (1995), in a similar analysis of British soccer teams from 1981 to 1991, find no statistically significant effect of playing on an artificial surface. Schwartz and Barsky (1977) argued anecdotally that if familiarity with the playing surface drove home-field advantage, we would expect to see the largest effects in baseball, since baseball fields are the most variable in their dimensions and playing surface.

The effect of rules has been the least-researched, perhaps because the rules of most sports confer no potential home advantage. Baseball is one of the few sports whose rules do indeed benefit the home team. Baseball grants home teams the privilege of batting last which affords an informational advantage. As an example, consider a team that is two runs behind in the 9th inning. Perhaps under normal circumstances a sacrifice play would make sense; however, because the team needs multiple runs in order to win the game, the manager chooses not to use an out. Softball, like baseball, grants the home team this informational advantage, and in a study of softball tournaments Courneya and Carron (1990) find that this rule has only a very small effect on outcomes. Thus, the home advantage in baseball does not seem to be rule-driven.

Having largely eliminated these three potential causes for home-field advantage, Nevill and Holder (1999), among other researchers, have concluded by default that fan support must be the primary driver of the home-field advantage. Some researchers have tried to determine whether larger crowds increase the home-field advantage; however, this research has been inconclusive. Pollard (1986) examines the four divisions of the English soccer league. He finds that attendance "density" (percent of capacity filled) varied widely among divisions, but that divisions with higher attendance did not have a greater home-field advantage.

Schwartz and Barsky (1977) study MLB and find a correlation between increased attendance and home team win percentage. They find that the home team wins 48% of the time when the crowd is small, 55% of the time when it is medium, and 57% of the time when it is large. Splitting our data this way yields similar results. Schwartz and Barsky recognize that it is possible that more fans do not improve a team's chance of winning, but rather that more fans come when their home team is playing well and is more likely to win. However they do not attempt to control for this endogeneity. Nor do they attempt to control for any of the other factors that affect a team's chance of winning—their win percentage at home, the win percentage of the away team on the road, etc. Schwartz and Barsky make no attempt to determine the marginal effect of an incremental fan on

* When this information was given to the Commission of Enquiry for the Football League, they decided to outlaw artificial fields for English and Scottish teams (Nevill and Holder 1999).

win percentage. In contrast, we adopt a rigorous approach that isolates the advantage conferred by home-field fans in baseball.

Prior findings of umpire discrimination motivate our explanation that higher attendance could generate umpire bias. Price and Wolfers (2007) find evidence of race discrimination by NBA referees. Bradbury (2007) shows racial discrimination is less likely to occur when the umpires are monitored by an electronic pitch tracking system called QuesTec. In a similar study, Parsons et. al (2007) show evidence of race discrimination among umpires. He finds that, indeed, the existence of QuesTec serves as a deterrent for racial discrimination. We employ a similar method to test the hypothesis that higher attendance leads to additional umpire bias thereby generating more home-field advantage.

If more fans help a team win, it is possible it would be profitable for teams to drop prices for the purpose of increasing the chance of victory, as the revenue associated with an additional win ($1.8M, as mentioned in the introduction) could outweigh lost ticket revenue. In this context it is important to understand consumers' elasticity of demand for attending MLB games. Many researchers have estimated MLB ticket demand elasticity, with a wide range of results. The most comprehensive analysis was done by Alexander (2001), who found elasticities ranging from -1.12 to -15.37 depending on city, with an average of -5.19. Older estimates have been lower. For example Scully (1989) estimated elasticity to be -0.63, Coffin (1996) estimated the range to be between -0.11 and -0.68, and Fort and Quick (1996) put the range to be between -0.14 and -0.36. However these older analyses treat price and attendance as exogenous factors, whereas Alexander's specification treats price and attendance as being endogenously determined by cost and attendance factors.

Data

Our game-level data comes from Retrosheet (www.retrosheet.org) for the years 1996 – 2005, a total of 28,096 observations. The data indicate time of day, attendance, ballpark, and home and away team statistics for each game. We also use Retrosheet to include the team's season record as a measure of team quality. We believe that season record is more indicative of quality of the team than the record to date because, at the beginning of a season, the small number of observations creates large variations in a team's record-to-date.

We supplement the game-level data with player-level data from the Baseball Archive (www.Baseball1.com). We include the starting pitchers' Earned Run Averages (ERAs) to control for the quality of home and away starting pitchers.

The game-level data include two potential instruments related to attendance—the day of the week that the game falls on and game time. We also acquired promotional data for games in 1996[†]. Promotions, such as a free bat or hat, are used to draw marginal fans for games that are predicted to be poorly attended. The promotional dataset includes a dummy variable for games with promotions and the cost paid by the team for the promotional gift, which serves as a proxy for the value of the promotion. Because the dataset is limited to games played in 1996, using promotions as an instrument limits our observations too severely to be effective; however, we do use the promotion data to approximate the elasticity of demand for baseball tickets to help evaluate implications for ticket pricing.

As an additional instrument, we use weather data from every city with a MLB team from 1996 – 2005. The data come from the National Climatic Data Center (http://www.ncdc.noaa.gov/oa/ncdc.html) and include daily values for mean temperature, maximum temperature, minimum temperature, precipitation amount, and dummy variables for occurrences of fog, rain or drizzle, snow or ice pellets, hail, thunder, or tornado. We create an "adverse weather" variable from these dummies, which equals one if adverse weather is present on a game day.

Finally, to test the theory that the umpire bias leads to additional home-field advantage, we implement a dummy variable for the existence of an electronic pitch monitoring system called QuesTec. The QuesTec system was installed in some ballparks in 2001 and 2002. While umpires still make calls during the game, the system monitors the accuracy of the umpire, thereby providing incentive to be unbiased. We use www.questec.com to identify the ballparks that have the QuesTec system.

Methodology

Specifications

To evaluate the impact of attendance on home team performance, we use a two-stage model, where performance is measured as the probability of a home win and attendance as attendance percentage of stadium capacity. To estimate the system we use instrumental variables (IV) to proxy for attendance. We use the IV approach to avoid potential endogeneity problems. For example, if it is the case that more fans attend a game when the home team has a greater chance of winning then we will find a positive correlation on attendance and home team performance even if greater attendance is not contributing to good performance. The Durbin-Wu-Hausman statistic confirms our intuition that we can reject exogeneity ($F(1,18434) = 20.59205$ corresponding with a p-value of $< .001$). The first stage of the 2SLS approach uses all exogenous predictors of attendance

[†] The promotional data were provided by Daniel Rascher, who used it to examine the economic effects of promotions in McDonald and Rascher (2000).

(controls from the primary regression and all valid instruments). Because everything included is exogenous, it is uncorrelated with the error term. In the second stage, we replace attendance with the approximated level of attendance from the first stage, thus allowing for a consistent estimator. While our IV approach allows us to correct for this endogeneity issue, it has limitations of its own – our model specification must not impose unintended moment restrictions. Therefore, we use an over-identification test to help validate our instrument choice. Additionally, our instruments must be strong predictors of attendance. We initially selected weather, time of game (weekend vs. weekday, day vs. night), and promotions as instruments, but found that some instruments worked better than others.[‡] The second-stage equation is:

$$\text{Performance} = \beta_0 + \beta_1\ \%\text{Attendance} + \beta_2\text{Controls} + \varepsilon \qquad [1]$$

We estimate this equation by two-stage least squares, using instruments. Thus the first-stage equation is:

[‡] In order for a variable to be considered a good instrument for attendance, it must meet two conditions: (1) it must be correlated with attendance and (2) it must be uncorrelated with the error term ε_i which includes any unobservables that might affect our dependent variable.

The first assumption can be tested using a first-stage F-Statistic, produced from the results of our first-stage regression of attendance on our instruments and controls.

Following our original specifications (Equation 1), we use ballpark and year fixed effects. We test the joint significance of the instruments using an F-Statistic as a measure of the strength of the correlation between our instruments and our treatment of attendance. We report the Wald F-statistic for weak instruments for our main regressions.

We test our second assumption using an over-identification test to examine the correlation between our instruments and the error term. We regress the residual (ehat) from our two-stage least squares on our fitted value of attendance from the first-stage regression, k-1 of our instruments (where k is the total number of instruments), and our controls:

$$\text{ehat}_{\text{iv}} = \beta_0 + \beta_1\text{Attendancehat} + \beta_2\text{Instruments} + \beta_3\text{Controls} + \varepsilon$$

We test the joint significance of our instruments as a measure of their correlation with the error term. We report the Hansen J-statistic test statistic for over-identification for our main regressions.

We can evaluate the effectiveness of our IV approach by examining whether two-stage least squares produces estimates that are statistically different from those found by OLS. The specification used in this test takes the following form:

$$\text{Performance} = \beta_0 + \beta_1\text{Attendancehat} + \beta_2\text{ehat}_{\text{first}} + \beta_3\text{Controls} + \varepsilon$$

Where Attendancehat is the fitted value of attendance produced by the first stage, and ehat_{iv} is the residual error term from the second stage. Our fixed effects are included here. We then test if the coefficient on Attendancehat (β_1) is equal to that on ehat_{iv} (β_2).

DOI: 10.2202/1559-0410.1192

$$\%\text{Attendance} = \beta_0 + \beta_1\text{Instruments} + \beta_2\text{Controls} + \varepsilon \qquad [2]$$

Our dependent variable %Attendance is calculated as (attendance/ballpark capacity) and controls for the fact that there is an upper limit on attendance. We considered several functional forms of the dependent variable including simply level of attendance. However, we speculated that using the level of attendance would result in bias when attendance levels near stadium capacity. We confirmed that this functional form does not satisfy normality assumptions of OLS. A censoring model would be difficult to implement as the upper bound, stadium capacity, varies by ballpark. Nevertheless, our results are robust, in both significance and magnitude, to several functional forms of the dependent variable including attendance level, log attendance, and log of percent attendance. Attendance measured as a percent of ballpark capacity is the most intuitive specification with normally distributed residuals.

In all regressions we use year and ballpark fixed effects. This allows us to control for other observable characteristics such as stadium capacity, field dimensions, and rosters, as well as unobservables, like psychological factors, that might be associated with a given ballpark[§].

Finally, we test the hypothesis that additional fans make the umpire more likely to make calls the home team. We do this by saving the fixed effect coefficients generated from our 2SLS model. There is a coefficient for each ballpark and year combination. We then check to see if the existence of QuesTec is a significant predictor of the magnitude of the attendance effect on score differential with the following model:

$$\text{Fixed Effect Coefficient} = \beta_0 + \beta_1\text{QuesTec Dummy} + \beta_2\text{Controls} + \varepsilon \; [3]$$

We control for an observation occurring for a QuesTec team or in a QuesTec year in case there are any characteristics of particular to the teams that purchased the QuesTec system or of the years when QuesTec is used. The standard errors are calculated for clustering by each ballpark and year combination.

[§] To test this specification, we run our basic regression using the instruments temperature, weekend game, and weekday game on the dependent variable, score differential, with varying levels of fixed effects. Our results are consistent with a number of different fixed effect specifications. We feel this specification is intuitive as attendance level is expected to depend on the popularity of teams. It also leaves us with enough in group heterogeneity to identify attendance effects.

Results

Descriptive Statistics

Table 1 shows seventeen offensive statistics, five pitching statistics, and six defensive statistics, listed for home and away teams. Tests of differences in means and medians show that many statistics vary significantly between the home and away teams. As expected, we see a statistically significant 6% home-field advantage in terms of winning percentage. Our measure of the difference is similar to the 7% difference measured by Carmichael and Thomas (2005). Potentially more interesting are the differences for specific offensive and defensive statistics. For example, the likelihood of being caught stealing a base is higher for the away team than for the home team. This difference could be because home team managers choose to steal less frequently, the home team runners are slightly faster, or because umpires are more likely to call the home team safe. The home team changes pitchers more frequently than the away team. Walks are more common for the home team and strikeouts more common for the away team, which could either be due to superior home team performance or umpire bias.

Table 2 provides descriptive statistics for all of the potential IVs. As the table shows, there are several viable candidates for suitable instrumental variables. In the next section, we discuss our motivation for selecting game time and temperature as our instruments.

Selecting Instruments

We intuitively specify game time as three categories (1) Weekend games: Friday night, Saturday and Sunday day and night games, (2) Weeknight games: Monday night, Tuesday night, Wednesday night, and Thursday night, and (3) Weekday games: Monday through Friday day games.[**]

[**] As mentioned in the Methodology section, in order for our instruments to be satisfactory they must meet two criteria: they must be correlated with attendance, and they must be uncorrelated with the LHS performance measure. To test the first criterion we calculate a Wald F-statistic. Our first stage produces an F-statistic of 765.13 with a probability of 0.0000, indicating that our chosen instruments are excellent predictors of attendance.

For our instruments to be valid, they must pass an Over-ID test as well as a test for weak instruments. We report a Kleibergen-Paap Wald F statistic and Hansen J-statistic for the regressions in Table 4. The K-P Wald F statistic is above the 5% critical values suggesting that we do not have a weak instrument problem in any of our estimates. The Hansen J-statistic of over-identification indicates that we can be confident our instruments pass the restrictions for over-identification.

DOI: 10.2202/1559-0410.1192

Table 1: Home and Away Team Game Level Summary Statistics.
Table 1 displays summary statistics and significance tests for difference of means and medians for various game level statistics for the home and away teams.

	Home Team			Away Team			P levels for test: home_stat = away_stat	
Statistic	Mean	Median	S.D.	Mean	Median	S.D.	T-test dif of means	Wilcoxon dif of medians
Game Statistics								
Win	0.53	1	0.5	0.47	0	0.5	0***	0***
Runs Scored	4.85	4	3.2	4.78	4	3.3	0.02**	0.02**
Season Record	0.5	0.5	0.1	0.5	0.5	0.1	0.45	0.29
Offensive Statistics								
At Bats	33.5	33	4.2	35.1	35	4.3	0***	0***
Plate Appearances	38.1	37	5.3	39.4	39	5.4	0***	0***
Strikeouts	6.17	6	2.7	6.72	7	2.8	0***	0***
Intentional Walks	0.27	0	0.6	0.26	0	0.6	0.11	0.06*
Walks	3.47	3	2.2	3.28	3	2.1	0***	0***
Awarded 1st	0.01	0	0.1	0	0	0.1	0***	0***
Hit by Pitch	0.35	0	0.6	0.35	0	0.6	0.76	0.37
Hits	9.02	9	3.4	9.21	9	3.6	0***	0***
Doubles	1.78	2	1.4	1.83	2	1.4	0***	0***
Triples	0.2	0	0.5	0.17	0	0.4	0***	0***
Home Runs	1.08	1	1.1	1.09	1	1.1	0.09*	0.17
Runners Batted In	4.61	4	3.1	4.54	4	3.2	0***	0***
Sacrificial Fly	0.28	0	0.5	0.28	0	0.5	0.96	0.79
Sacrificial Hit	0.35	0	0.6	0.34	0	0.6	0***	0***
Stolen Bases	0.61	0	0.9	0.61	0	0.9	0.4	0.82
Caught Stealing	0.25	0	0.5	0.28	0	0.5	0***	0***
Grounded to Double Play	0.78	1	0.9	0.8	1	0.9	0***	0***
Players Left on Base	7.1	7	2.7	7.11	7	2.7	0.76	0.22
Defensive Statistics								
Putouts	27.5	27	2.2	26	27	2.7	0***	0***
Assists	10.5	10	3.3	9.97	10	3.2	0***	0***
Double Play	0.97	1	1	0.93	1	0.9	0***	0***
Triple Play	0	0	0	0	0	0	0.51	0.51
Errors	0.68	0	0.9	0.67	0	0.9	0.64	0.26
Passed Balls	0.09	0	0.3	0.07	0	0.3	0***	0***
Pitching Statistics								
Starting Pitcher ERA	4.66	4.49	1.5	4.66	4.49	1.5	0.63	0.63
Number of Pitchers	3.65	4	1.3	3.54	3	1.3	0***	0***
Balks	0.04	0	0.2	0.04	0	0.2	0***	0***
Wild Pitch	0.31	0	0.6	0.31	0	0.6	0.5	0.19
Individual Earned Runs	4.37	4	3.1	4.47	4	3.1	0***	0***
Team Earned Runs	4.37	4	3.1	4.46	4	3.1	0***	0***

*** dif in home and away stats are significant at the 1% level, ** at the 5% level, * at the 10% level

Table 2: Game Attendance and Instrumental Variable Summary Statistics. Table 2 reports the summary statistics for the instrumental variable data. Percent attendance is equal to (attendance/stadium capacity). Adverse weather is a dummy variable equal to one for the existence of any adverse weather condition including rain, hail, snow and poor visibility. Game time variables are dummies that equal one if the game occurred during that time. Promotional game is a dummy variable that equals one if there was a promotion during the game (i.e. free bat day). Value of a promotion is the cost of a promotional give away per person per game.

	Mean	Std. Dev.	Min	Max	No. Obs.
Game Statistics:					
Attendance	29674.95	11637.32	746	61707	27883
Percent Attendance	0.60	0.24	0.01	1	27883
Game Length in Minutes	174.58	26.88	79	395	27658
Instrumental Variable Statistics:					
Temperature	69.03	10.70	15.4	103	22949
Temp max	81.12	11.36	34	116.1	13476
Temp min	59.45	11.17	18	91	13629
Adverse Weather	0.38	0.49	0	1	22949
Night Game	0.66	0.47	0	1	27884
Weeknight Game	0.48	0.50	0	1	27884
Weekday Game	0.12	0.32	0	1	27884
Weekend Game	0.40	0.49	0	1	27884
Friday	0.16	0.36	0	1	27884
Saturday	0.17	0.38	0	1	27884
Sunday	0.16	0.37	0	1	27884
Monday	0.10	0.30	0	1	27884
Tuesday	0.14	0.34	0	1	27884
Wednesday	0.16	0.36	0	1	27884
Thursday	0.11	0.32	0	1	27884
Promotional Game	0.24	0.43	0	1	2097
Value of Promotion	0.20	0.71	0	4.95	2097

We find adverse weather to be an unsatisfactory instrument. Including adverse weather produced results that implied a significant, positive effect of attendance on performance, similar in magnitude to that found in our final specification, it performed poorly in the over-ID test. Adverse weather showed signs of being positively correlated with the run differential and likelihood of a home win. Adverse weather is a dummy which equals one for any day with rain, snow, hail, or fog but can be interpreted as rain in most cases during the baseball season. When controlling for home team or ballpark fixed effects, this bias could potentially result from a home team advantage of playing in rain in a home stadium. Also, on rainy days the umpire decides when to cancel the game. If umpire bias exists, then umpires might allow games to continue longer if the home team is down a few runs and call games sooner when the home team is winning. We therefore dropped our adverse weather variable. We kept temperature as an instrument, as it proves effective when controlling for either the home team or ballpark. [††]

Score Differential and Home Win as Dependent Variables

We run our chosen specification (ballpark and year fixed effects) with two LHS measures of performance—home win and score differential. The first stage results for run differential are shown in Table 3. Visiting team record is the strongest determinant of attendance; with a ten percentage point increase in the visiting team's record corresponding to about a 3% increase in percent attendance (i.e. if a given team's attendance were 50% of capacity, increasing the visiting team's record by 10% would result in attendance moving up to 53% capacity). Temperature also has a significant impact; a ten degree increase in temperature leads to a 3% increase in percent attendance. Weekend games tend to average ballparks that are 12% more full than weeknight games.

The second stage results answer the question that we initially posed—do fans help the home team win? Indeed, we find a significant value for attendance on score differential, with a coefficient of 2.62 on percent attendance, as shown in the first column of Table 4b. This coefficient indicates that increasing attendance by one standard deviation (about 25%) would result in approximately .64 additional runs for the home team. The results for likelihood of a home win are shown in the third column of Table 4. We use a probit model so determining the marginal effects of higher attendance is somewhat calculation intensive. However, the coefficients imply that increasing percent attendance one standard deviation from the mean (going from 60% to 85% attendance) increases the likelihood of a home team win by 5.4%. Increasing percent attendance by one standard deviation around the mean (going from 48% to 73% attendance) increases the likelihood of a home team win by 5.5%.

[††] Additionally, we determined that while variables describing promotional games could be a good instrument, there are many missing observations. Because we have promotion data for only 1996, when using the promotions we lose approximately 21,000 observations, rendering our coefficients insignificant. It is interesting to note, however, that the results from our first stage regression with promotions data indicate that running a promotion results on average in a 16.7% increase in attendance for that game. These findings are in line with those found by Rascher and McDonald (2000).

Table 3: Varying Fixed Effects Specifications. Table 3 displays the results of the first stage of our 2SLS model where the first stage predicts attendance with the equation: %Attendance = β_0 + β_1Instruments + β_2Controls + ε. As shown, the regression explains much of the variation in percent attendance. Our three instrumental variables (temperature, weekend game, and weekday game) are all significant at the 1% level. Robust t-statistics are reported in parentheses.

	Percent Attendance
Temperature	0.00264***
	(0.000142)
Weekend Game	0.133***
	(0.00227)
Weekday Game	0.0452***
	(0.00386)
Rain Dummy	-0.00469*
	(0.00255)
Rain Dummy*Dome Dummy	0.0122**
	(0.00480)
Domed Stadium Dummy	-0.000117
	(0.00300)
Home Average to Date	0.140***
	(0.0200)
Visiting Season Record	0.324***
	(0.0245)
Visiting Average to Date	-0.00275
	(0.0174)
Home Pitcher ERA	-0.00478***
	(0.000829)
Visiting Pitcher ERA	0.00329***
	(0.000912)
Constant	0.133***
	(0.0211)
Number of Observations	18711
R-squared	0.759
F Statistic	665.1

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Examining Alternative Dependent Variables

We next examine possible causes for the attendance effect on home-field by testing the changes in other performance measures caused by attendance. The results are shown in Table 5. Testing the attendance effect for other game statistics provides a better picture as to why the significant coefficients on run

Table 4: The Attendance Effect on Score Differential and Home Win Likelihood. Table 4 reports the results of the 2S IV model where the first stage predicts attendance with the equation: %Attendance = $\beta_0 + \beta_1$Instruments + β_2Controls + ε. The second stage uses the fitted values for %Attendance to predict the causal effects of attendance on run differential using: Performance = $\beta_0 + \beta_1$ % Attendance + β_2Controls + ε. We use 2SLS for the score differential Statistic and 2S probit model for home win likelihood. We also report the corresponding OLS and probit models respectively. Robust t-statistics are reported in parentheses. As shown, percent attendance has a significant impact on the run differential and the likelihood of a home win.

	Score Difference				Home Win Likelihood			
	2SLS Second Stage		OLS		2S Probit Second Stage		Probit	
Percent Attendance	2.615***		0.598**		0.559***		-0.0128	
	(0.55)		(0.242)		(0.178)		(-0.0822)	
Rain Dummy	-0.243***		-0.285***		-0.021		-0.0228	
	(-0.0887)		(-0.0885)		(-0.029)		(-0.0287)	
Domed Stadium Dummy	-0.173*		-0.177*		-0.0059		-0.00869	
	(-0.104)		(-0.105)		(-0.0339)		(-0.0334)	
Rain Dummy * Dome Dummy	0.667***		0.715***		0.041		0.058	
	(0.168)		(0.168)		(0.055)		(0.0545)	
Home Record to Date	5.894***		6.154***		3.759***		3.652***	
	(0.469)		(0.467)		(0.222)		(0.209)	
Visiting Season Record	-3.487***		-2.863***		-0.106		0.0286	
	(-0.746)		(-0.73)		(-0.257)		(0.25)	
Visiting Record to Date	-4.692***		-4.671***		-2.737***		-2.766***	
	(-0.457)		(-0.459)		(-0.178)		(-0.177)	
Home Pitcher ERA	-0.313***		-0.321***		-0.0503***		-0.0608***	
	(-0.0277)		(-0.0281)		(-0.00857)		(-0.00872)	
Visiting Pitcher ERA	0.386***		0.393***		0.0819***		0.0829***	
	(0.0449)		(0.0462)		(0.0133)		(0.0134)	
Constant	-1.510***		-0.607		-1.231***		-0.792***	
	(-0.571)		(-0.539)		(-0.276)		(-0.186)	
	Obs	18711	Obs	18711	Obs	18412	Obs	18711
	2nd Stg R-sqr	0.282	R-sqr	0.285	Wald chi-sqr	15.06	Ps. R-sqr	0.218
	2nd Stg F	33.77	F-Statistic	32.77	chi-sqr p	0.0001	Chi-Square	4614
	Over ID (Hansen J-stat)	1.561						

Table 5: Attendance Effects on other Game Statistics. Table 5 displays the second stage results from the model specified in Table 4 while varying the dependent variable. The dependent variable is the difference (statistic$_{\text{Home}}$ – statistic$_{\text{Away}}$) for a variety of game statistics. The coefficient on attendance indicates the attendance effect on the dependent statistic. For example, the coefficient on strikeout means that higher home team attendance leads to a significantly lower average number of strike-outs for the home team. All hitting statistics (the first six columns) are scaled by plate appearances. There are approximately 39 plate appearances per team per game. Robust t-statistics are reported in parentheses.

	Strikeouts	Walks	Hits	Doubles	Triples	Home Runs	Stolen Base Rate	Number of Pitchers	ERA
Percent Attendence	-0.0834***	-0.00229	0.0690***	0.0425***	-0.00468**	0.0117***	-0.139	0.0489	-2.308***
	(0.0120)	(0.00835)	(0.0125)	(0.00580)	(0.00191)	(0.00452)	(0.109)	(0.180)	(0.527)
Rain Dummy	-.00693***	0.00178	-.00579***	-.00256***	4.82e-05	4.46e-05	-0.0310*	-0.0611**	0.342***
	(0.00195)	(0.00132)	(0.00210)	(0.000973)	(0.000310)	(0.000716)	(0.0186)	(0.0298)	(0.0847)
Domed Stadium	-0.000593	0.00336**	-0.00198	-0.00115	-0.00109***	0.00139	-0.0692***	0.0268	0.196*
Dummy	(0.00231)	(0.00152)	(0.00234)	(0.00105)	(0.000362)	(0.000874)	(0.0223)	(0.0341)	(0.100)
Rain Dummy*Dome	0.00361	-0.0100***	0.0102***	0.00209	-0.000415	0.000148	0.0809**	0.0164	-0.503***
Dummy	(0.00368)	(0.00253)	(0.00373)	(0.00173)	(0.000581)	(0.00138)	(0.0350)	(0.0544)	(0.159)
Home Record to	-0.103***	0.0371***	0.0761***	-0.00584	0.00486***	0.0227***	0.374***	-1.092***	-5.728***
Date	(0.0119)	(0.00811)	(0.0113)	(0.00537)	(0.00171)	(0.00426)	(0.111)	(0.163)	(0.447)
Visiting Season	0.177***	-0.0988***	-0.0347**	-0.0210***	0.0138***	-0.0358***	0.269	0.350	2.708***
Record	(0.0165)	(0.0111)	(0.0163)	(0.00796)	(0.00251)	(0.00601)	(0.167)	(0.239)	(0.704)
Visiting Record to	-0.0341***	-0.0242***	-0.0421***	-0.0295***	-0.00525***	-0.0104***	-0.0692	0.0790	4.349***
Date	(0.0108)	(0.00747)	(0.0107)	(0.00513)	(0.00163)	(0.00384)	(0.106)	(0.154)	(0.433)
Home Pitcher ERA	0.00706***	-.00396***	-.00414***	-.00229***	-.000372***	-.00105***	-0.000665	0.102***	0.288***
	(0.000745)	(0.000388)	(0.000572)	(0.000295)	(8.61e-05)	(0.000228)	(0.00564)	(0.00933)	(0.0267)
Visiting Pitcher ERA	-.00589***	0.00388***	0.00417***	0.00201***	0.000459***	0.00120***	0.0207***	-0.125***	-0.344***
	(0.000659)	(0.000405)	(0.000673)	(0.000400)	(9.52e-05)	(0.000249)	(0.00603)	(0.0111)	(0.0402)
Constant	0.0603***	0.0664***	-0.100***	-0.00410	0.00403**	-0.0142***	0.353***	0.863***	1.277**
	(0.0126)	(0.00814)	(0.0127)	(0.00593)	(0.00188)	(0.00450)	(0.129)	(0.187)	(0.545)
Obs	18711	18711	18711	18711	18711	18711	8638	18711	18711
2nd Stg R-sqr	0.265	0.311	0.239	0.215	0.215	0.206	0.242	0.203	0.273
2nd Stg F	34.12	48.67	25.74	25.02	19.31	21.06	404.2	23.48	32.21
Over ID J-stat	0.901	14.96	34.99	76.78	45.80	11.10	8.544	122.2	3.727
Chi-sq P-value	0.637	0.000563	2.53e-08	0	1.13e-10	0.00389	0.0140	0	0.155

Robust standard errors in parentheses: *** p<0.01, ** p<0.05, * p<0.1

differential and home win likelihood exist. For example, when scaling the hitting statistics by plate appearances we find that the home team strikes out less frequently when attendance is higher. Assuming the same number of plate appearances for the home and away team (we assume 39), if attendance were to increase by one standard deviation we would expect the home team to strikeout .79 fewer times.

We also see more hitting success with higher attendance. When scaling the hitting statistics by plate appearances, the coefficient on attendance when regressed on home-away team hits indicates that it would require a 37% increase to generate an additional hit. The coefficients on doubles requires that attendance increase by 60% to generate an additional double. Increasing attendance by 50% would result in .23 more home runs for the home team. While these effects are small, attendance in baseball does have a high variance, and the results indicate that higher attendance can give the home team a real hitting advantage.

We also see that pitching success is affected by higher attendance. A decrease in one earned run would require a 43% increase in attendance. One explanation is that pitcher's performances are affected by higher attendance.

Umpire Bias Results

It is possible that the size of the home crowd influences calls made by umpires. As shown, higher attendance leads to both fewer strikeouts and a reduced home team earned run average. Both of these statistics are influenced directly by the decisions of umpires. While attendance has an effect on all several offensive categories, not just those directly influenced by umpires, bias towards the home team in calling strikes will lead to more favorable pitch counts, more runners on base, and more at-bats for the home team, all factors that would lead to an improvement in other offensive categories.

We hypothesize that if higher attendance leads to additional umpire bias, then lower levels of home-field advantage will arise when a QuesTec system exists. If it is more costly for the umpires to gratify the fans, they will do it less frequently. We test this with a regression of the ballpark and year fixed effect coefficient generated from our 2SLS model on a dummy variable that indicates the existence of a QuesTec system as shown in equation 3 of the methodology section. The results of these regressions are shown in table 6 for several offensive statistics. As shown, we fail to reject the hypothesis that attendance generated home-field advantage is lower in stadiums with QuesTec than those without QuesTec.

Table 6. Questec's Effect on Attendance Effects on Performance. Table 6 displays the results of the regression Ballpark_Year_Coefficient = $\beta_0 + \beta_1$QuesTec_Dummy + β_2Controls + ε. The ballpark_year coefficient is generated from the ballpark_year fixed effect coefficient in our 2SLS model where the first stage predicts attendance with the equation: %Attendance = $\beta_0 + \beta_1$Instruments + β_2Controls + ε. The second stage uses the fitted values for %Attendance to predict the causal effects of attendance on run differential using: Performance = $\beta_0 + \beta_1$ % Attendance + β_2Controls + ε. Robust t-statistics are reported in parentheses. As shown, the existence of QuesTec does not impact the effect of attendance on various performance statistics.

	Score Differential	Home Win Likelihood	Strikeouts	Walks	Hits	Home Runs	Stolen Base Rate	Earned Run Average
QuesTec Dummy	0.522	0.155	-0.00500	-0.00756	0.0110	0.00495	0.000920	-0.536
	(0.439)	(0.144)	(0.0111)	(0.00723)	(0.0111)	(0.00356)	(0.0622)	(0.422)
QuesTec Year Dummy	-0.587**	-0.164**	0.00986	0.00216	-0.0106*	-0.00218	0.0399	0.619**
	(0.253)	(0.0830)	(0.00640)	(0.00417)	(0.00638)	(0.00205)	(0.0354)	(0.243)
QuesTec Team Dummy	-0.104	-0.0670	-0.00236	0.00208	-0.00120	-0.000708	-0.0186	0.0246
	(0.293)	(0.0962)	(0.00741)	(0.00483)	(0.00740)	(0.00238)	(0.0410)	(0.282)
Constant	0.656***	0.254***	-0.0419***	-0.0181***	0.0644***	0.0183***	-0.151***	-0.329**
	(0.157)	(0.0517)	(0.00398)	(0.00260)	(0.00397)	(0.00128)	(0.0219)	(0.151)
Observations	255	255	255	255	255	255	249	255
R-squared	0.022	0.015	0.012	0.005	0.012	0.011	0.009	0.028
F-Statistic	1.866	1.300	1.052	0.414	1.026	0.905	0.714	2.416

Standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Implications for Ticket Pricing

At the outset of this paper we suggested that a positive coefficient on attendance may mean ticket prices are optimal at a level lower than they would be if attendance had no effect. While we do not have sufficient ticket price data to do a full analysis of optimal ticket pricing we can make some conclusions based on previous estimates of demand elasticity in baseball. Table 7 reports the necessary decrease in prices in order to generate a one standard deviation move in attendance for three price brackets and four values of price elasticity ranging from -10 to -.1 based on the measurements of Alexander (2001), Scully (1989), Coffin (1996), and Fort and Quick (1996). The table assumes a stadium capacity of 50 thousand. As shown, a very conservative elasticity of -1.5 results in "break-even" ticket revenues. That is, the increase in the quantity of sales exactly offsets the decreased revenue in price changes given our assumptions[‡‡].

Table 7. Ticket Price Analysis using Elasticity of Demand. Table 7 reports the necessary ticket price decreases in order to generate a one standard deviation increase in attendance using price elasticity of demand estimates consistent with the previous literature. It also reports the new ticket revenue resulting from the increase of sales due to the price decrease. As shown a conservative estimate of elasticity, -1.5, will result in a break even price change to generate a one standard deviation move in attendance. The table assumes a stadium capacity of 50 thousand and that the move in attendance is around the mean.

Price decrease necessary to generate a 1 standard deviation change in attendance			
	Starting Price of Ticket		
Price Elasticity of Demand	25	50	100
-10	-1.25	-2.5	-5
-5	-2.5	-5	-10
-1.5	-8.33	-16.67	-33.33
-1	-12.5	-25	-50
Change in ticket revenues resulting from decreased prices and increased demand (000s)			
	Starting Price of Ticket		
Price Elasticity of Demand	25	50	100
-10	255	510	1020
-5	165	330	660
-1.5	0	0	0
-1	-285	-570	-1140

Because a one standard deviation increase in attendance generates a four percent increase in the likelihood of a home win, it's possible that decreasing

‡‡ These figures doesn't take into account increased concession sales.

ticket prices could result in additional revenue from a more successful season record. In our one standard deviation move example, we would expect an additional $72,000 of revenue using the estimate of Silver (2006) that an additional win generates $1.8 million in revenue. The use of very conservative figures shows potential for modest profit increases and reveals that our findings have sufficient magnitude to suggest that further research on optimal ticket pricing would be valuable[§§].

Conclusions

In our analysis of home-field advantage in baseball, we find that increased attendance has a significant effect on many game statistics, ultimately leading to an increased likelihood of winning for the home team. We find the instrumental variable approach successfully controls for potential endogeneity problems. Temperature and game time are excellent instruments. They show a strong correlation with attendance and are not found to be correlated with the team performance. These results are intuitive, as neither time nor temperature is expected to give one team a systematic advantage. Rain and promotions are poor instruments. Both do show a sufficiently strong correlation with attendance, but rain shows some correlation with home team performance, and using promotional data decreases our dataset by over 90%, pushing the standard errors to a level where we lose explanatory power.

We find a significant positive effect of attendance on score differential, likelihood of a home win and almost all other major offensive categories. Likewise, we find a significant negative effect of attendance on home team ERA, indicating that higher attendance causes home pitchers to have greater success. Because we were unable to find supporting evidence for an attendance effect on umpire bias, we conclude that attendance increases home win percentage via increased home team performance (or diminished away team performance). We find that for a 38% increase in attendance (i.e. an increase in attendance from 40% of stadium capacity to 78% of stadium capacity) we should expect the home team to score one additional run. Baseball teams rarely experience swings in attendance this large, however it is important to remember that a single run in baseball is tremendously meaningful—in 2006, MLB teams averaged 4.86 runs per game, so adding an additional run represents a 20.6% increase in scoring. This result is confirmed by our estimate that a one standard deviation increase in attendance increases the home win likelihood with by four percent. These results indicate that smaller, realistic variations in home attendance have a meaningful impact on a game's outcome.

[§§] This analysis is only relevant for stadiums that have sufficiently low attendance to be able to generate significantly higher attendance levels using just price decreases.

DOI: 10.2202/1559-0410.1192

Our findings give some indications as to the mechanisms of causality by which the attendance effect works. Intuitively it can be argued that greater attendance might give a psychological edge to the home team, but such a claim is extremely hard to test. As mentioned above, it is possible that attendance affects the lineup and substitution choices made by the home manager. Our results indicate a positive correlation between pitching substitutions and attendance, which supports this hypothesis. However, because we also see a negative correlation between earned runs allowed and attendance, this substitution effect may simply be a consequence of pitchers performing more poorly when they have been pitching for several innings. While a change in managerial decisions may be the cause of additional home-field advantage, we see significant increases in performance for several measures that would be unaffected by manager choices. It is also intuitively difficult to argue that managers would make these types of game winning choices only when attendance is high. Therefore, our findings suggest that managerial choices are not the primary cause for attendance based home-field advantage.

We were also unable to find evidence of the competing hypothesis that any additional advantage is a product of umpire bias as opposed to performance increases. Our methodology mirrored that of Bradbury (2007) who found that umpire racial discrimination decreased in QuesTec ballparks. We found no such indication of umpire behavioral changes with increases in attendance. We therefore conclude that fans do matter and can increase baseball players' ability to perform.

Finally, it is worth noting that our results suggest that team records have greater dispersion than inherent team quality. Being good/bad in MLB is a virtuous/vicious cycle. Good teams attract more fans, which in turn further increase that team's chances of winning. Bad teams, on the other hand, attract very few fans, which makes it even more difficult than it otherwise would for them to win. Therefore, it is reasonable that a team could reach a "tipping point"—if a bad team begins to win more games, then the concomitant increase in attendance will further aid them in the future.[***]

[***] We believe that a more robust dataset could add further precision to our conclusions. MLB promotions statistics are largely unavailable to the public; we were able to procure data only for the 1996 season. Using this data limits our number observations to a point where it is impossible to draw meaningful conclusions. Still, we believe that promotions have value as an instrument, and more extensive promotion data would be useful in further isolating the effect of attendance.

We should point out that we only used the QuesTec data to determine if umpire bias *increases* as a result of additional attendance (i.e. does higher attendance lead to more biased calls?). We do not establish if there is any non-attendance-related bias in the first place, as this would be tangential to our study. It would be interesting to use the QuesTec data in a study similar to that of Bradbury (2007) to determine if umpire bias is less likely to exist in ballparks with QuesTec.

References

Acee, K. (2002, June 21). Bolts Hope to Price Out Raiders Fans with Policy. *San Diego Union-Tribune*.

Alexander, D. (2001). Major League Baseball: Monopoly Pricing and Profit-Maximizing Behavior. *Journal of Sports Economics*, 2, 341-355.

Barnett, V. and Hilditch, S. (1993). The Effect of an Artificial Pitch Surface on Home Team Performance in Football. *Journal of the Royal Statistical Society*, 156, 39-50.

Blau, J. (2006, February 9). Fans Stand Behind Team During Difficult Season. *The Penn State Daily Collegian*.

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

Bradbury, J.C., *The Baseball Economist: The Real Game Exposed*. New York: Dutton, 2007.

Carmichael, F. and Thomas, D. (2005). Home-Field Advantage and Team Performance: Evidence for English Premiership Football. *Journal of Sports Economics*, 6, 164-281.

Clarke, S. and Norman, J. (1995). Home Ground Advantage of Individual Clubs in English Soccer. *Statistician*, 44, 509-521.

Coffin, D. (1996). If you build it will they come? Attendance and new stadium construction. In J. Fizel, E. Gustafson, & L. Hadley (Eds.), *Baseball economics* (pp. 33-46). Westport, CT: Praeger.

Courneya, K. and Carron, A. (1990). Batting First Versus Last: Implications for the Home Advantage. *Journal of Sport & Exercise Psychology*, 12, 312-316.

Courneya, K. and Carron, A. (1991). Effects of Travel and Length of Home Stand/Road Trip on the Home Advantage. *Journal of Sport & Exercise Psychology,* 13, 42-49.

Courneya, K. and Carron, A. (1992). The Home Advantage in Sport Competitions. *Journal of Sport & Exercise Psychology,* 14, 13-27.

DOI: 10.2202/1559-0410.1192

Demmert, H. (1973). *The Economics of Professional Sports*. Lexington, MA: Lexington Books.

Fort, R., and Quirk, J. (1995). Cross-subsidization, Incentives, and Outcomes in Professional Sports. *Journal of Economic Literature*, 33, 1265-1299.

McDonald, M. and Rascher, D. (2000). Does bat day make cents?: The effect of Promotions on the Demand for Major League Baseball. *Journal of Sport Management,* 14, 8-27.

Nevill, A. and Holder, R. (1999). Home Advantage in Sport. *Sports Medicine,* 28, 221-236.

Pace, A. and Carron, A. (1992). Travel and the Home Advantage. *Canadian Journal of Sport Science,* 17, 60-64.

Parsons, C.A., Sulaeman, J., Yates, M. & Hamermesh, D. (2007). Strike Three: Umpires' Demand for Discrimination. *NBER Working Paper Series*, Working Paper 13665.

Pollard, R. (1986). Home Field Advantage in Soccer. *Journal of Sports Science,* 4, 237-248.

Price, J. and Wolfers, J. (2007). Racial Discrimination Among NBA Referees. *NBER Working Paper Series*, Working Paper 13206.

Ramano, J. (2001, February 27). Baseball Adapts to a New Zone. *St. Petersburg Times.*

Schwartz, B. and Barsky, S. (1977). The Home Advantage. *Social Forces,* 55, 641-661.

Scully, G. (1989). *The Business of Major League Baseball*. Chicago: University of Chicago Press.

Silver, N. (2006). Is Alex Rodriguez Overpaid?. In J. Keri (Ed.), *Baseball by the Numbers* (174-98). New York: Basic Books.

Welki, A. and Zlatoper, T. (1994). US Professional Football: The Demand for Game-Day Attendance in 1991. *Managerial and Decision Economics*, 15, 489-495.