Predicting Day-to-Day Variability in Baseball Attendance to Support Staffing

Eric Webb

December 15, 2014

Abstract

Attendance at Major League Baseball games can fluctuate greatly from game to game. Service organizations in the surrounding area rely upon an accurate forecast of fan attendance at baseball games in order to appropriately staff their services. Over-staffing leads to excess labor costs while under-staffing leads to lost sales and unhappy fans. This paper uses 30 years of attendance data in fitting an econometric model to forecast fan attendance, noting that certain model covariates become known to schedule planners at different times. While the away team and the time of the game are known well in advance, the weather and the recent performance of the home team are not known until just before the game. By examining the relative effects of different covariates, we make suggestions for when planners should forecast attendance and schedule workers and when planners should augment or decrease staffing based on unforseen events.

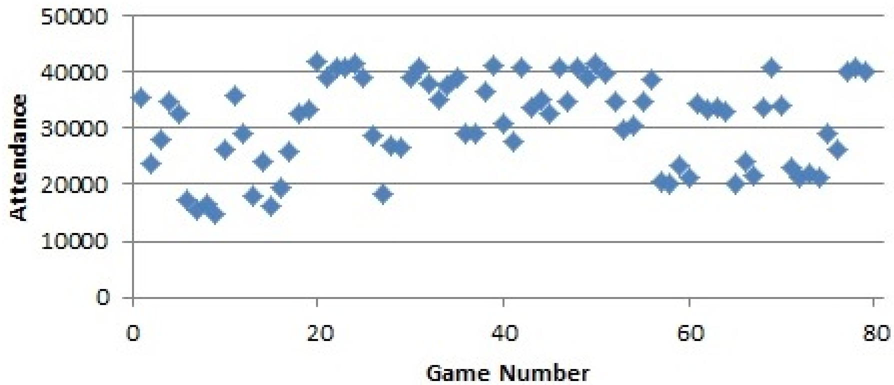
1 Introduction

Baseball attendance has the potential to fluctuate greatly from game-to-game. Total attendance at Major League Baseball (MLB) games exceeded 74 million in 2013, an average of over 30,000 a game (Baseball-Reference.com 2013). All MLB stadiums had a capacity of at least 37,000 in 2013 (ballparksofbaseball.com 2014), and most games did not sell enough tickets to reach the stadium's capacity. Baseball attendance tends to be more variable than other professional sports in the United States. Attendance can be a function of the day of the week, the time of the game, the quality of the opponent, the current record of the home team, the weather, and many other variables.

An example of attendance fluctuation is given in Figure 1, where the home attendance of the author's favorite team, the Cincinnati Reds, is graphed for 2013. Home game attendance ranged from 14,916 to 43,168 for the Reds, exhibiting significant game-to-game variation.

Around each MLB stadium, dozens of service organizations rely upon the patronage of baseball fans. The more fans that attend a game, the more customers these businesses will receive on game day. Such businesses include restaurants, bars, hotels, and other attractions. As more customers request service, more staff members need to be scheduled to work to provide adequate service capacity. If not enough workers are scheduled, the business may miss out on service opportunities due to inadequate capacity. However, if

Figure 1: Attendance at Cincinnati Reds home games in 2013.



too many workers are scheduled, the business will have idle workers and excessive labor costs. Thus, an accurate forecast of attendance is the first step in accurately staffing the businesses surrounding a baseball stadium.

This paper will seek to develop a forecasting model for baseball attendance and establish staffing recommendations for these businesses that cater to baseball fans.

Staffing service organizations in light of uncertain demand is a common theme in operations literature. What makes staffing to serve baseball fans more challenging is the multitude and variety of possible factors affecting attendance. Factors related to the date and time of the game, the weather, the attractiveness of the matchup, the home team's performance, the away team's performance, and the pitching matchup can all be relevant to a fan's decision of whether to attend the game or not. These factors become known at different times. The date, time, and opponent are available well in advance of the game, but the weather and the pitching matchup may not be known until a few days before the game. Additionally, the overall performance and playoff chances of the home team are constantly changing. In late August, a restaurant might forecast higher attendance for a team in September that is in the midst of a chase for a playoff spot. But if the team starts losing and falls out of the race in early September, the attendance forecast will turn out to be overly optimistic as ticket sales drop off.

In this paper, the author assumes that there is no information sharing between the baseball team and the surrounding businesses. Specifically, it is assumed that the surrounding businesses are not made aware of current ticket sales data. Such data would simplify the task of forecasting total attendance to simply forecasting additional tickets sales between the time the data was made available and the start of the game. There would be less uncertainty in such a simpler forecast. The operations staff of the baseball team will rely on such ticket sales information when staffing ushers, concession vendors, parking attendants, security, and merchandise sales people at the stadium. The surrounding businesses, not being privy to such information, will need to rely upon an econometric forecast of attendance. Thus, the task of developing a schedule for staffing at a huge baseball stadium may actually be simpler than scheduling staff at the smaller surrounding businesses.

The forecasting model presented in this paper is built upon 30 years of attendance data. The model suggests that the effect of the covariates which are known well in advance of game day, including the timing of the game, the opponent, and performance in previous years, are more significant in the attendance forecast than the covariates which are unknown until the week of the game, including the recent performance of the home team and the weather. The best-fitting model for the data is provided, showing the effect on attendance for each covariate. Businesses in the ecosystem surrounding baseball stadiums are encouraged to schedule their staff as far in advance as desired, typically 2-4 weeks, using their best guess as to the value of the weather and performance covariates. As the day of the game nears and the exact values of these covariates are revealed, the business can update its staffing if the initial forecast was too far off. This "forecast early and update" plan of action seems to be more appropriate than a "wait and see" approach that waits until all covariates are fully revealed to create a schedule.

The rest of the paper is organized as follows: Section 2 discusses relevant literature related to forecasting baseball attendance and staffing in the face of uncertain demand. Section 3 introduces the data used in fitting the model of Section 4. Section 5 gives the results of fitting the model to the available data. Section 6 discusses future work and concludes.

2 Literature Review

Relevant literature falls into two groups: past efforts to forecast sporting attendance and staff scheduling literature.

2.1 Sports Attendance Forecasting Literature

Many papers have attempted to build forecasting models to predict game-day attendance. In baseball, Hill et al. (1982) and Lemke et al. (2010) use a single year of data to build a forecasting model to predict game-to-game fluctuations in attendance, while Beckman et al. (2011) uses multiple years. Donihue et al. (2007) looks at spring training baseball games. Other papers have examined attendance forecasting in other sports, including soccer (Forrest and Simmons (2002), García and Rodríguez (2002), Madalozzo and Villar (2009)), Australian rules football (Borland and Lye (1992)), U.S. football (Welki and Zlatoper (1994) and Welki and Zlatoper (1999)), and basketball (Zhang et al. (1995)).

Additionally, many papers have looked at the effect of specific factors in predicting baseball attendance. Kahane and Shmanske (1997) examines the effect of roster turnover on fan attendance. McDonald and Rascher (2000) finds that promotions increase attendance, but that having too many promotions lessens each promotion's effect. Player strikes lessen attendance in the year after the strike, according to Schmidt and Berri (2004), but do not have a long-term impact. Whitney (1988) looks at how fans respond to the probability that their team will win the World Series. Butler (2002) looks at the effect of interleague matchups on attendance. Papers such as Knowles et al. (1992) and Meehan et al. (2007) have examined how fan attendance varies with the probability the home team will win.

2.2 Staff Scheduling Literature

The operations literature has focused in depth upon staffing decisions when facing demand uncertainty. Once a schedule has been created and staffing forecasted, staffing adjustments can be costly. Thompson (1999) gives a method for altering workforce schedules in real time. Hur et al. (2004) and Mehrotra et al. (2010) look at the tradeoff between schedule stability and profitability. Van Mieghem (2003) provides an overview of capacity adjustments under uncertain demand.

Much of the literature also focuses upon the potential benefits of using cross-trained workers when demand is uncertain. Cross-trained workers tend to be more flexible, filling any gaps in staffing, while also costing more in salary and sacrificing a bit in efficiency. Pinker and Shumsky (2000) examines the tradeoff between service quality and service efficiency and Easton (2014) examines the benefit of cross-trained workers in the case of demand uncertainty and uncertain worker attendance.

3 Data

The data for this project comes from Baseball-Reference.com. Details about every Major League Baseball game from 1980-2013 were extracted. In order to use three years of lagged performance covariates, only the data from 1983-2013 were used in the analysis. This data includes 69,497 games with an average attendance of 28,487 per game. There were 26 teams in MLB in 1983, with 2 expansion teams being added in 1993 and 2 more expansion teams in 1998. Each team is scheduled to play 162 games per season, but the full 162 games are not always played if a game is cancelled due to weather and the game's result would not change playoff determinations.

3.1 Covariates Used in Model

The covariates used in analysis can be split in two ways. First, certain covariates are constant for a given team in a given year. These covariates are marked "constant". An example would be whether or not the team made the playoffs last year. Other covariates have the ability to change throughout the season. These are marked "changing" and an example would be the current place in division for the home team.

Another way to split data is based on when the value for the covariate is known. Covariates known well in advance of the game, such as the date and time of the game, are marked "advance". Covariates that become known the week of the game, including the current games back of the division leader, are marked "last minute". These two groupings of covariates will be useful in determining what is important to focus upon for staffing considerations, as a business would like to know when it is reasonable to put out a work schedule for its employees.

Covariates derived from the Baseball-Reference.com data include:

- stad_und_2: Indicator that the home stadium was built within the last 2 years. Constant, advance.
- stad_und_5: Indicator that the home stadium was built within the last 5 years. Constant, advance.
- stad_und_10: Indicator that the home stadium was built within the last 10 years. Constant, advance.
- playoffs_yr3: Indictor that the home team made the playoffs 3 years ago. Constant, advance.

- playoffs_yr2: Indictor that the home team made the playoffs 2 years ago. Constant, advance.
- playoffs_yr1: Indictor that the home team made the playoffs last year. Constant, advance.
- ws_yr3: Indicator that the home team made the World Series 3 years ago. Constant, advance.
- ws_yr2: Indicator that the home team made the World Series 2 years ago. Constant, advance.
- ws_yr1: Indicator that the home team made the World Series last year. Constant, advance.
- april: Indicator that the game was played in April. Changing, advance.
- may: Indicator that the game was played in May. Changing, advance.
- june: Indicator that the game was played in June. Changing, advance.
- july: Indicator that the game was played in July. Changing, advance.
- august: Indicator that the game was played in August. Changing, advance.
- september: Indicator that the game was played in September. Changing, advance.
- october: Indicator that the game was played in October. Changing, advance.
- mon_day: Indicator that the game was played on Monday during the day. Changing, advance.
- mon_night: Indicator that the game was played on Monday at night. Changing, advance.
- tues_day: Indicator that the game was played on Tuesday during the day. Changing, advance.
- tues_night: Indicator that the game was played on Tuesday at night. Changing, advance.
- wed_day: Indicator that the game was played on Wednesday during the day. Changing, advance.
- wed_night: Indicator that the game was played on Wednesday at night. Changing, advance.
- thur_day: Indicator that the game was played on Thursday during the day. Changing, advance.
- thur_night: Indicator that the game was played on Thursday at night. Changing, advance.
- fri_day: Indicator that the game was played on Friday during the day. Changing, advance.
- fri_night: Indicator that the game was played on Friday at night. Changing, advance.
- sat_day: Indicator that the game was played on Saturday during the day. Changing, advance.
- sat_night: Indicator that the game was played on Saturday at night. Changing, advance.
- sun_day: Indicator that the game was played on Sunday during the day. Changing, advance.
- sun_night: Indicator that the game was played on Sunday at night. Changing, advance.
- in_division: Indicator that the home and away teams are in the same division. Changing, advance.
- interleague: Indicator that the home and away teams are in different leagues. Changing, advance.
- nyy_away: Indicator that the New York Yankees are the away team. Changing, advance.
- bos_away: Indicator that the Boston Red Sox are the away team. Changing, advance.
- chc_away: Indicator that the Chicago Cubs are the away team. Changing, advance.
- opening_day: Indicator that the game is the first home game of the season for the home team. Changing, Advance.
- doubleheader: Indicator that two games are to be played on the same day. Changing, last minute.
- rain: Indicator that rain is known to have fallen during the game. Changing, last minute.
- no_rain: Indicator that no known rain fell during the game. Changing, last minute.
- win_perc_3yr: The percentage of games won by the home team 3 years ago. Constant, advance.

- win_perc_2yr: The percentage of games won by the home team 2 years ago. Constant, advance.
- win_perc_1yr: The percentage of games won by the home team last year. Constant, advance.
- place_div_3: The season-ending place in the division for the home team 3 years ago. If the team won its division, the value is 1. If it finished in 2nd, the value is 2. Etc. Constant, advance.
- place_div_2: The season-ending place in the division for the home team 2 years ago. Constant, advance.
- place_div_1: The season-ending place in the division for the home team last year. Constant, advance.
- place_in_div: The current place in the division for the home team. Changing, last minute.
- games_back: The current number of games behind the division leader for the home team. If in first place, the value is 0. Changing, last minute.
- games_up: The current number of games ahead of the 2nd place team in division for the home team. If not in first place, the value is 0. Changing, last minute.
- eliminated: The minimum of 1 and games_back divided by the number of games remaining. Represents a proxy for playoff probability. When eliminated is 1, the home team cannot win its division. Changing, last minute.
- wins_last_10: The number of wins in the last 10 games for the home team. Changing, last minute.
- runs_last_10: The number of runs in the last 10 games for the home team. Changing, last minute.
- temperature: The temperature at the start of the game. Changing, last minute.
- streak: If the home team won its last game, this is the number of games won in a row. Otherwise, it is the negative of the number of games lost in a row (i.e. 3 losses in a row would be -3). Changing, last minute.

Some covariates merit further discussion about their construction or significance:

The variables "rain" and "no_rain" are derived from the weather description in the data as follows: If "rain", "snow", "drizzle", or "showers" is present in the weather description, rain is set to 1 and no_rain is 0. If a weather description is not present or is "Unknown", both rain and no_rain are set to 0. In all other cases, no_rain is set to 1 while rain is 0. Upon examination of all possible weather descriptions, this seems to adequately describe the data available. Approximately 80% of the data have either rain or no_rain set to 1.

Doubleheaders are typically scheduled in order to make up a game that was rained out. Sometimes the doubleheader is scheduled for the next day and sometimes it is scheduled for months in the future. Thus, the value of "doubleheader" becomes known at different points depending on the doubleheader. Here it is marked as "last minute" for simplicity.

Over the years, the division structure and the playoff structure of MLB have changed. Despite expanded playoffs and smaller divisions in recent years, the definitions pertaining to a playoff appearance and to the place in division will not be changed over time in the data. If a team played any extra games beyond the initial 162 scheduled games, then it is assumed to have made the playoffs. Whether there were 4 teams or 7 teams in the division of the home team, the place in division is always taken to be the ranking of the team relative to its division mates.

There are two leagues in MLB: the National League and the American League. The winner of each league meets to play in the World Series. Prior to 1997, teams from different leagues never met during the season. Interleague play began that year and continues to this day. A portion of the schedule is now dedicated to playing teams from the opposite league.

The New York Yankees, Boston Red Sox, and Chicago Cubs tend to draw more fans when they are the visiting team than other visiting teams. Thus the indicators on the Yankees, Red Sox, and Cubs as the visiting team are included.

Figure 2 shows summary statistics for each covariate over the 69,497 games included in the data set. The statistics have been divided into the dependent variable (attendance), explanatory variables that are indicators, and explanatory variables that are quantitative.

One issue with the data is that the recorded attendance represents the number of tickets sold, not the number of people who actually attended the game. While nearby businesses would be more interested in the number of people physically present at the game, this information is usually unavailable. Number of tickets sold is the best proxy available and is typically a good approximation of the actual attendance. This number will be the dependent variable in the forecast.

3.2 Missing Covariates

For a multi-year analysis such as this, data on certain attendance covariates, used in previous one-year analyses such as Hill et al. (1982) and Lemke et al. (2010), are difficult to compile and will therefore be excluded from analysis. Data on promotions and post-game fireworks is unavailable. Details about star players, current injuries, and national or local TV broadcasts are similarly unavailable in useable form. Details specific to a particular team's city in a particular year, such as population, unemployment, and per capita income, are not analyzed. The expected win probability, available for calculation via published betting odds, is not used because historic betting data is unavailable. Certain other small variables used in other analyses, such as the minority status of the starting pitchers or an indicator for the public school districts in the home city being on vacation, are unavailable and not used.

The starting pitchers of the home and away team are available in the data, but time constraints have forced their exclusion from analysis for the time being. They will be included in the analysis in the future to determine the effect on attendance from successful star pitchers.

Ticket pricing information will not be used in analysis for two reasons. The first is practical: it is difficult to get accurate historic pricing information. The second reason is that Fort (2004) found that teams tend to price tickets in the inelastic portion of the demand curve, minimizing any possible effect of small price changes.

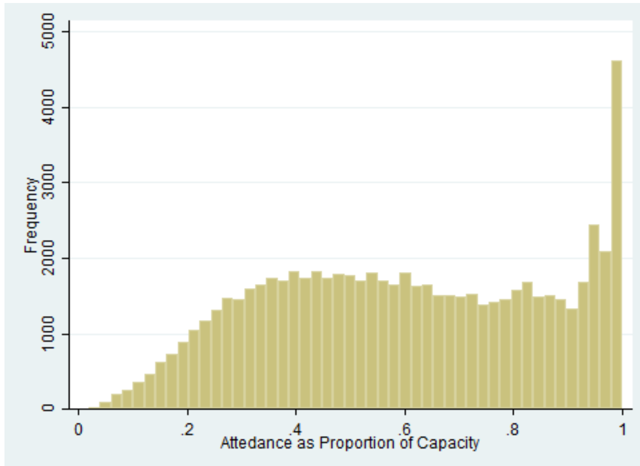
4 Model

We wish to use all available covariates that were as described in Section 3 to fit a model of attendance. It is expected that new stadiums (stad_under_2, stad_under_5, stad_under_10), recent playoff success (playoffs_yr3, playoffs_yr2, playoffs_yr1, ws_yr3, ws_yr2, ws_yr1), summer months (june, july, august), weekend

Figure 2: Summary Statistics for covariates included in the model.

Variable	Obs	Mean	Std. Dev.	Min	Max
DEPENDENT VARIABLE					
attendance	69497	28487.32	11874.38	746	80227
INDICATOR VARIABLES					
stad_und_2	69497	.0597148	.2369594	0	1
stad_und_5	69497	.1442221	.3513173	0	1
stad_und_10	69497	.2779832	.4480083	0	1
playoffs_yr3	69497	.2156352	.4112652	0	1
playoffs_yr2	69497	.2202685	.4144306	0	1
playoffs_yr1	69497	.2231463	.4163587	0	1
ws_yr3	69497	.0683051	.2522706	0	1
ws_yr2	69497	.0684058	.2524428	0	1
ws_yr1	69497	.0686648	.2528851	0	1
april	69497	.1416176	.34866	0	1
may	69497	.1727557	.3780386	0	1
june	69497	.1701944	.3758062	0	1
july	69497	.1662518	.3723092	0	1
august	69497	.1741946	.3792794	0	1
september	69497	.1621077	.3685522	0	1
october	69497	.0121012	.1093388	0	1
mon_day	69497	.0170079	.1293016	0	1
mon_night	69497	.0901622	.286416	0	1
tues_day	69497	.0095544	.0972791	0	1
tues_night	69497	.1402363	.3472345	0	1
wed_day	69497	.033282	.1793733	0	1
wed_night	69497	.1185087	.3232119	0	1
thur_day	69497	.0388794	.1933088	0	1
thur_night	69497	.0720319	.2585426	0	1
fri_day	69497	.0086047	.0923622	0	1
fri_night	69497	.1502367	.3573059	0	1
sat_day	69497	.0667079	.2495173	0	1
sat_night	69497	.0942055	.2921165	0	1
sun_day	69497	.1465531	.3536624	0	1
sun_night	69497	.0140294	.1176127	0	1
in_division	69497	.4394434	.4963229	0	1
interleague	69497	.0612257	.2397455	0	1
nyy_away	69497	.035354	.1846744	0	1
bos_away	69497	.035095	.1840214	0	1
chc_away	69497	.0351958	.1842756	0	1
opening_day	69497	.0126624	.1118135	0	1
doubleheader	69497	.0120725	.1092103	0	1
rain	69497	.0231952	.1505243	0	1
no_rain	69497	.7788538	.4150218	0	1
QUANTITATIVE VARIABLES					
win_perc_3yr	69497	.5000771	.0681061	.2654321	.7160494
win_perc_2yr	69497	.5000931	.0681611	.2654321	.7160494
win_perc_1yr	69497	.500096	.0681135	.2654321	.7160494
place_div_3	69497	3.280271	1.671169	1	7
place_div_2	69497	3.258731	1.660033	1	7
place_div_1	69497	3.239003	1.64792	1	7
place_in_div	69497	3.167547	1.654999	1	7
games_back	69497	7.466963	7.831034	0	52
games_up	69497	.7060808	2.23103	0	29
eliminated	69497	.1906648	.2923714	0	1
wins_last_10	69497	4.978215	1.663022	0	10
runs_last_10	69497	45.40321	11.2182	13	98
temperature	69497	73.10314	9.678068	12	109
streak	69497	.2310747	2.540543	-15	20

Figure 3: The attendance is censored at stadium capacity. This histogram shows that a significant number of games are censored, rendering the Ordinary Least Squares estimate of the model coefficients most likely biased.



games (fri_night, sat_day, sat_night, sun_day, sun_night), in_division games, interleague games, popular opponents (nyy_away, bos_away, chc_away), opening_day, doubleheader, no_rain, strong past performance (win_perc_3yr, win_perc_2yr, win_perc_1yr), games_up, wins_last_10, runs_last_10, temperature, and streak will all be positively related to attendance. Non-summer months (april, may, september, october), weekday games (mon_day, mon_night, tues_day, tues_night, wed_day, wed_night, thur_day, thur_night, fri_day), rain, place in division (place_div_3, place_div_2, place_div_1, place_in_div), games_back, and eliminated are all expected to be negatively related to attendance.

Initially, attendance will be forecasted via an Ordinary Least Squares (OLS) regression. All covariates will be used. Additionally, a fixed effect will be included for each year used in the regression (1983-2013) and for each of the 30 baseball teams. Each team will keep the same fixed effect for all years, even if it changed its home city or team name.

Because the model relies upon collinear indicator variables, a reference point will be set in the regression, thereby dropping certain indicator covariates from the regression. The signs of the remaining covariates will be relative to that reference point. The reference point in this model will be a home game of the Cincinnati Reds played on a Sunday night in March in 1983 with both rain and no_rain set to 0.

Results of the OLS regression are shown in Section 5.

OLS returns biased coefficients if the dependent variable is censored. Here, attendance is censored if the stadium sells all tickets and reaches its capacity. Even if more people were interested in buying tickets, they would not be allowed to and the attendance would remain at the level of the stadium capacity. Figure 3 shows a histogram of attendance as a percentage of the stadium capacity. As can be seen, quite a few games sell out of tickets and reach capacity, censoring the attendance. Thus, the OLS coefficients may be biased in this regression.

To eliminate the bias of OLS, the model will also be run with a Tobit Regression. Tobit Regression has a similar interpretation to OLS while estimating coefficients with Maximum Likelihood Estimation and allowing for the dependent variable to be censored. In Stata, a single censoring point must be given for all data points. In reality, each data point is potentially censored at the stadium capacity. The stadium capacity varies from game to game. So a single censoring point cannot be given for the attendance dependent variable. As a work-around, we will instead use the dependent variable “perc_full”, which is defined as the attendance divided by the stadium capacity. If the game sold out of tickets, perc_full is 1.0. Thus, perc_full can be uniformly censored at 1.0 across all games. The result of the Tobit Regression of perc_full on the covariates of Section 3 is given in Section 5.

5 Results

5.1 OLS Model

The results of the OLS regression are given in Table 1. The reference fixed effects are the year 1983 and the team Cincinnati Reds. The monthly effects are compared to a game in March and the daily effects are compared to a game on a Sunday night. The dip in attendance in 1995 is due to a player strike, as discussed in Schmidt and Berri (2004).

The coefficients of all covariates except streak, which has a small magnitude, are in the direction predicted in Section 4. All covariates are significant except win_perc_3yr, may, fri_night, and rain. Thus, Friday night and Sunday night attract a similar number of fans, while March and May games attract a similar number of fans.

At first glance, it is surprising that the rain covariate is not significant. Fans do not like to get wet, so it was suspected that this covariate would be much more meaningful. However, it may be that there is a selection bias in the games that are played. If it rains too heavily, the game will be cancelled and removed from the data set. So those games in which it rained do not represent all days in which a game was scheduled and it rained. After removing rained-out games, only 2% of games listed rain in their weather description. Fans may also respond to rain forecasts more than actual, experienced rain. So if a weather forecast calls for rain and no rain occurs, fans may stay away from the ballpark despite the weather description of the game not listing any rain during the game. For this reason, it may be more meaningful in the future to include a 24 hour advance weather forecast as the covariate instead of the actual rainfall during the game.

The OLS regression had an R^2 of .626, suggesting 62.6% of the variation in attendance from game to game can be explained by the regression covariates. A variance inflation test was run on the covariates to determine if any were unduly influencing the standard errors of the covariates. None of the covariates had VIF higher than 10 except for the month indicators. It was later realized that the reference month, March, only had 54 games out of the possible 69,497. By making March the reference month, the other monthly indicators summed to one for over 99.9% of all games, making the other months collinear. By switching the reference month to April in a later regression (not shown), the problem was solved. The VIF of the monthly indicators dropped to a reasonable level while the coefficients of the regression remained almost entirely unchanged.

Table 1: Ordinary Least Squares Model Fit and Regression Output

Source	SS	df	MS
Model	6.1293e+12	111	5.5219e+10
Residual	3.6697e+12	69385	52889596.9
Total	9.7990e+12	69496	141000927

Number of obs	= 69497
F(111, 69385)	= 1044.03
Prob > F	= 0.0000
R-squared	= 0.6255
Adj R-squared	= 0.6249
Root MSE	= 7272.5

attendance	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
team						
PIT	-2054.791	212.4924	-9.67	0.000	-2471.275	-1638.306
STL	6958.748	211.0073	32.98	0.000	6545.174	7372.322
MIL	1585.725	211.1063	7.51	0.000	1171.957	1999.493
CHC	11044.13	222.9626	49.53	0.000	10607.13	11481.14
ATL	-448.3953	214.9156	-2.09	0.037	-869.6295	-27.16103
PHI	3899.266	210.4134	18.53	0.000	3486.856	4311.676
WSN	-3762.164	209.7688	-17.93	0.000	-4173.311	-3351.018
MIA	-6273.586	237.2663	-26.44	0.000	-6738.628	-5808.545
NYM	6104.773	210.5813	28.99	0.000	5692.034	6517.512
COL	11176.79	234.6264	47.64	0.000	10716.92	11636.66
SDP	1156.643	209.4312	5.52	0.000	746.1582	1567.128
SFG	2710.654	211.5373	12.81	0.000	2296.041	3125.267
LAD	14182.38	211.7541	66.98	0.000	13767.34	14597.41
ARI	1006.368	253.5696	3.97	0.000	509.372	1503.364
BOS	4649.44	216.24	21.50	0.000	4225.61	5073.27
NYY	5807.957	219.7917	26.42	0.000	5377.165	6238.748
BAL	7988.743	211.5754	37.76	0.000	7574.056	8403.431
TOR	4355.115	210.9328	20.65	0.000	3941.687	4768.543
TBR	-7485.049	258.3017	-28.98	0.000	-7991.32	-6978.778
CLE	-382.4557	214.1907	-1.79	0.074	-802.2692	37.35777
CHW	-2231.561	210.5073	-10.60	0.000	-2644.155	-1818.967
DET	541.7912	211.0492	2.57	0.010	128.1351	955.4473
KCR	957.9108	213.5113	4.49	0.000	539.429	1376.393
MIN	-2124.736	211.6854	-10.04	0.000	-2539.639	-1709.833

TEX	2355.339	212.7035	11.07	0.000	1938.441	2772.237
HOU	37.08223	209.4004	0.18	0.859	-373.3421	447.5066
SEA	1830.739	212.9238	8.60	0.000	1413.408	2248.069
OAK	-3413.574	214.9356	-15.88	0.000	-3834.847	-2992.301
LAA	6691.766	211.5336	31.63	0.000	6277.161	7106.371
year						
1984	191.0754	231.7753	0.82	0.410	-263.2038	645.3546
1985	1063.611	230.7823	4.61	0.000	611.2782	1515.944
1986	1400.729	230.9661	6.06	0.000	948.0354	1853.422
1987	3747.612	230.3233	16.27	0.000	3296.179	4199.046
1988	4401.38	246.0763	17.89	0.000	3919.071	4883.689
1989	5032.665	230.9318	21.79	0.000	4580.039	5485.291
1990	4940.01	237.5684	20.79	0.000	4474.376	5405.643
1991	4841.784	255.5456	18.95	0.000	4340.915	5342.652
1992	4589.557	255.9006	17.93	0.000	4087.992	5091.122
1993	8903.122	238.5688	37.32	0.000	8435.528	9370.717
1994	7788.522	262.2912	29.69	0.000	7274.432	8302.612
1995	1192.364	254.9638	4.68	0.000	692.6354	1692.093
1996	3554.27	256.7516	13.84	0.000	3051.037	4057.503
1997	4029.057	256.21	15.73	0.000	3526.885	4531.228
1998	5353.251	258.3782	20.72	0.000	4846.83	5859.672
1999	4895.371	255.8962	19.13	0.000	4393.815	5396.927
2000	5000.653	254.5072	19.65	0.000	4501.819	5499.486
2001	4582.986	254.6191	18.00	0.000	4083.933	5082.039
2002	3445.484	254.8174	13.52	0.000	2946.042	3944.926
2003	3476.859	254.4193	13.67	0.000	2978.197	3975.52
2004	5201.955	254.3373	20.45	0.000	4703.454	5700.455
2005	6311.153	253.175	24.93	0.000	5814.931	6807.376
2006	6829.857	253.7986	26.91	0.000	6332.412	7327.301
2007	8187.088	253.4155	32.31	0.000	7690.394	8683.782
2008	8264.668	253.173	32.64	0.000	7768.449	8760.886
2009	5625.39	252.9245	22.24	0.000	5129.658	6121.121
2010	5589.308	252.6404	22.12	0.000	5094.133	6084.483
2011	6609.724	252.7617	26.15	0.000	6114.312	7105.137
2012	6707.283	252.8101	26.53	0.000	6211.776	7202.791
2013	6696.168	253.1471	26.45	0.000	6200	7192.336

stadium_under_2	4324.459	151.1323	28.61	0.000	4028.24	4620.678
stadium_under_5	2107.186	124.2264	16.96	0.000	1863.702	2350.669
stadium_under_10	4050.881	90.91813	44.56	0.000	3872.681	4229.08
win_perc_3yr	463.7853	881.5899	0.53	0.599	-1264.129	2191.7
win_perc_2yr	6237.931	901.5542	6.92	0.000	4470.887	8004.976
win_perc_1yr	24402.13	892.1901	27.35	0.000	22653.44	26150.83
place_in_div_yr3	-233.7358	36.50622	-6.40	0.000	-305.2879	-162.1837
place_in_div_yr2	-135.2875	36.77499	-3.68	0.000	-207.3665	-63.20863
place_in_div_yr1	-131.5508	37.12358	-3.54	0.000	-204.3129	-58.78865
playoffs_yr3	899.4774	104.2902	8.62	0.000	695.0688	1103.886
playoffs_yr2	1359.813	103.4365	13.15	0.000	1157.077	1562.548
playoffs_yr1	760.2291	104.6362	7.27	0.000	555.1423	965.3159
ws_yr3	683.9985	131.5767	5.20	0.000	426.1083	941.8887
ws_yr2	1868.407	130.4501	14.32	0.000	1612.725	2124.088
ws_yr1	2483.441	129.4949	19.18	0.000	2229.631	2737.25
april	-2122.668	1021.709	-2.08	0.038	-4125.215	-120.1197
may	143.1666	1026.994	0.14	0.889	-1869.74	2156.073
june	2466.884	1028.444	2.40	0.016	451.1367	4482.632
july	4287.759	1028.547	4.17	0.000	2271.809	6303.708
august	4112.536	1028.925	4.00	0.000	2095.845	6129.226
september	2109.794	1032.745	2.04	0.041	85.6157	4133.972
october	2679.814	1065.38	2.52	0.012	591.6706	4767.958
mon_day	-2299.725	323.1807	-7.12	0.000	-2933.158	-1666.291
mon_night	-4794.173	252.1225	-19.02	0.000	-5288.332	-4300.013
tues_day	-5717.718	373.5535	-15.31	0.000	-6449.882	-4985.554
tues_night	-4950.591	246.1958	-20.11	0.000	-5433.134	-4468.047
wed_day	-4048.581	280.8372	-14.42	0.000	-4599.022	-3498.141
wed_night	-4627.869	248.2437	-18.64	0.000	-5114.427	-4141.312
thur_day	-3743.483	274.6635	-13.63	0.000	-4281.823	-3205.143
thur_night	-4582.527	256.2983	-17.88	0.000	-5084.871	-4080.183
fri_day	-2495.678	396.5427	-6.29	0.000	-3272.901	-1718.455
fri_night	285.2907	245.498	1.16	0.245	-195.8849	766.4662
sat_day	2377.364	259.3271	9.17	0.000	1869.084	2885.645
sat_night	4766.645	251.5689	18.95	0.000	4273.571	5259.72
sun_day	1572.169	246.1202	6.39	0.000	1089.774	2054.564
sun_night	0	(omitted)				

in_division	926.4062	60.39497	15.34	0.000	808.0322	1044.78
interleague	1925.989	132.2361	14.56	0.000	1666.806	2185.171
nyy_away	6361.96	152.7024	41.66	0.000	6062.663	6661.256
bos_away	3644.306	153.1539	23.80	0.000	3344.124	3944.487
chc_away	2970.2	152.5846	19.47	0.000	2671.134	3269.265
opening_day	19855.88	282.7519	70.22	0.000	19301.69	20410.08
place_in_div	-716.3618	27.73117	-25.83	0.000	-770.7148	-662.0087
games_back	-118.8356	8.560986	-13.88	0.000	-135.6151	-102.0561
games_up	177.1362	14.91008	11.88	0.000	147.9125	206.3599
games_back_over_games_remaining	-2954.416	237.7291	-12.43	0.000	-3420.364	-2488.467
doubleheader	2168.751	254.9911	8.51	0.000	1668.969	2668.533
wins_last_10	126.465	22.55776	5.61	0.000	82.2518	170.6781
runs_scored_last_10	6.371788	3.159729	2.02	0.044	.1787244	12.56485
rain	-390.8676	225.0656	-1.74	0.082	-831.9957	50.26053
no_rain	507.2379	140.7373	3.60	0.000	231.3931	783.0828
temperature	55.18673	3.607298	15.30	0.000	48.11643	62.25703
streak	-27.13349	11.54463	-2.35	0.019	-49.76095	-4.506032
_cons	1167.492	1374.479	0.85	0.396	-1526.484	3861.468

5.2 Tobit Model

The results of the Tobit regression are shown in Table 2. The reference fixed effects are still 1983 and the Cincinnati Reds. The monthly effects are still compared to a game in March and the daily effects are compared to a game on a Sunday night.

The Tobit model seems to fit the data slightly better than the OLS model. 64.2% of the variation in perc_full from game to game is explained by the regression covariates, slightly higher than the 62.6% of the OLS model. The conclusions of the effects of each covariate are largely the same. In the Tobit model, win_perc_3yr, may, june, september, october, fri_night, runs_scored_last_10, rain, and streak are not statistically significant at the .05 level. All other covariates are significant and in the direction expected.

Table 2: Tobit Model Fit and Regression Output

note: sun_night omitted because of collinearity

Tobit regression	Number of obs	=	69497
	LR chi2(111)	=	70928.13
	Prob > chi2	=	0.0000
Log likelihood = 26776.405	Pseudo R2	=	4.0821

perc_full	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
team_index						
PIT	.0184338	.0045632	4.04	0.000	.0094898	.0273777
STL	.1356865	.0045366	29.91	0.000	.1267947	.1445782
MIL	.0593472	.00455	13.04	0.000	.0504292	.0682651
CHC	.3813795	.0048091	79.30	0.000	.3719537	.3908053
ATL	-.0374751	.0046185	-8.11	0.000	-.0465274	-.0284228
PHI	.0440345	.004585	9.60	0.000	.035048	.053021
WSN	-.0446686	.0045046	-9.92	0.000	-.0534976	-.0358396
MIA	-.0206606	.0051099	-4.04	0.000	-.030676	-.0106451
NYM	.0653906	.0045198	14.47	0.000	.0565317	.0742494
COL	.142935	.0050356	28.38	0.000	.1330652	.1528048
SDP	.0071048	.0045024	1.58	0.115	-.0017199	.0159295
SFG	.0901868	.0045857	19.67	0.000	.0811988	.0991747
LAD	.2029625	.0045481	44.63	0.000	.1940483	.2118767
ARI	-.0265869	.0054437	-4.88	0.000	-.0372566	-.0159173
BOS	.257932	.0046423	55.56	0.000	.248833	.267031
NYY	.021493	.0047209	4.55	0.000	.01224	.0307459
BAL	.1554404	.0045457	34.20	0.000	.1465309	.1643499
TOR	.0745242	.0045387	16.42	0.000	.0656283	.0834201
TBR	-.1505038	.0055448	-27.14	0.000	-.1613716	-.1396361
CLE	.0035332	.004598	0.77	0.442	-.005479	.0125453
CHW	.0235325	.0045279	5.20	0.000	.0146579	.0324071
DET	.0552325	.0045548	12.13	0.000	.046305	.0641599
KCR	.1323109	.0045842	28.86	0.000	.1233258	.1412959
MIN	-.0905392	.004544	-19.93	0.000	-.0994454	-.081633

TEX	.06449	.0045689	14.12	0.000	.055535	.073445
HOU	.0269838	.0045096	5.98	0.000	.0181451	.0358226
SEA	.0002066	.004571	0.05	0.964	-.0087525	.0091657
OAK	.1296224	.0046505	27.87	0.000	.1205074	.1387374
LAA	.209495	.0045443	46.10	0.000	.2005881	.2184019
year						
1984	.0052804	.0049723	1.06	0.288	-.0044654	.0150262
1985	.0212817	.004952	4.30	0.000	.0115759	.0309876
1986	.027438	.0049558	5.54	0.000	.0177248	.0371513
1987	.0835845	.0049427	16.91	0.000	.0738967	.0932722
1988	.09711	.0052845	18.38	0.000	.0867524	.1074676
1989	.1037814	.0049592	20.93	0.000	.0940612	.1135015
1990	.1044493	.0051055	20.46	0.000	.0944424	.1144561
1991	.1053962	.0054954	19.18	0.000	.0946254	.1161671
1992	.1024521	.0054978	18.64	0.000	.0916764	.1132278
1993	.173125	.0051369	33.70	0.000	.1630568	.1831933
1994	.1501717	.0056406	26.62	0.000	.1391161	.1612273
1995	.0292437	.0054733	5.34	0.000	.0185161	.0399713
1996	.0744001	.0055132	13.50	0.000	.0635944	.0852059
1997	.0821474	.005502	14.93	0.000	.0713634	.0929313
1998	.1133402	.0055499	20.42	0.000	.1024624	.124218
1999	.104178	.0054957	18.96	0.000	.0934065	.1149495
2000	.1212417	.0054685	22.17	0.000	.1105235	.1319598
2001	.1204032	.0054752	21.99	0.000	.1096718	.1311346
2002	.0919719	.0054755	16.80	0.000	.08124	.1027039
2003	.1004063	.0054725	18.35	0.000	.0896801	.1111325
2004	.1538102	.0054764	28.09	0.000	.1430763	.164544
2005	.1771314	.0054459	32.53	0.000	.1664575	.1878053
2006	.1892251	.0054593	34.66	0.000	.1785249	.1999253
2007	.2239141	.0054636	40.98	0.000	.2132054	.2346228
2008	.2286738	.0054594	41.89	0.000	.2179733	.2393742
2009	.1756958	.0054463	32.26	0.000	.165021	.1863706
2010	.1855227	.0054412	34.10	0.000	.174858	.1961874
2011	.2153945	.0054507	39.52	0.000	.2047111	.2260778
2012	.2146234	.0054459	39.41	0.000	.2039494	.2252974
2013	.2126103	.0054466	39.04	0.000	.201935	.2232855

stadium_under_2	.0904881	.00327	27.67	0.000	.0840788	.0968973
stadium_under_5	.0561515	.0026932	20.85	0.000	.0508728	.0614303
stadium_under_10	.1473481	.0019677	74.88	0.000	.1434915	.1512048
win_perc_3yr	.0080525	.0189894	0.42	0.672	-.0291668	.0452717
win_perc_2yr	.1504581	.0194081	7.75	0.000	.1124183	.1884978
win_perc_1yr	.5418237	.0192111	28.20	0.000	.50417	.5794774
place_in_div_yr3	-.0060209	.0007857	-7.66	0.000	-.0075608	-.004481
place_in_div_yr2	-.003456	.000792	-4.36	0.000	-.0050082	-.0019037
place_in_div_yr1	-.0037739	.0007992	-4.72	0.000	-.0053402	-.0022076
playoffs_yr3	.0103314	.0022503	4.59	0.000	.0059209	.0147419
playoffs_yr2	.0176813	.0022317	7.92	0.000	.0133071	.0220555
playoffs_yr1	.0077128	.0022573	3.42	0.001	.0032885	.0121371
ws_yr3	.0238841	.0028459	8.39	0.000	.0183061	.0294621
ws_yr2	.0536032	.0028367	18.90	0.000	.0480433	.0591631
ws_yr1	.0680803	.0028239	24.11	0.000	.0625455	.0736152
april	-.0687517	.0244016	-2.82	0.005	-.1165787	-.0209247
may	-.0177103	.0245036	-0.72	0.470	-.0657373	.0303167
june	.0321177	.0245321	1.31	0.190	-.0159651	.0802005
july	.0724819	.0245342	2.95	0.003	.0243948	.1205689
august	.0700545	.024542	2.85	0.004	.0219523	.1181567
september	.0280197	.0246172	1.14	0.255	-.02023	.0762694
october	.0417151	.025256	1.65	0.099	-.0077867	.0912169
mon_day	-.0466262	.007002	-6.66	0.000	-.06035	-.0329023
mon_night	-.0999316	.0054456	-18.35	0.000	-.110605	-.0892582
tues_day	-.1335672	.008064	-16.56	0.000	-.1493726	-.1177618
tues_night	-.1051125	.0053187	-19.76	0.000	-.1155371	-.0946879
wed_day	-.0889905	.0060615	-14.68	0.000	-.100871	-.0771101
wed_night	-.0968395	.0053623	-18.06	0.000	-.1073497	-.0863293
thur_day	-.0815411	.0059319	-13.75	0.000	-.0931676	-.0699146
thur_night	-.0954594	.0055343	-17.25	0.000	-.1063066	-.0846122
fri_day	-.0500828	.0086351	-5.80	0.000	-.0670076	-.0331581
fri_night	.004871	.0053053	0.92	0.359	-.0055274	.0152693
sat_day	.0496951	.0056087	8.86	0.000	.038702	.0606882
sat_night	.1024357	.0054387	18.83	0.000	.0917758	.1130956
sun_day	.0305242	.0053193	5.74	0.000	.0200984	.04095
sun_night	0	(omitted)				

```
                  in_division |   .0191362   .0013014    14.70   0.000     .0165855    .0216869
                  interleague |   .0449128   .0028617    15.69   0.000     .0393039    .0505217
                     nyy_away |    .136244   .0033072    41.20   0.000     .1297618    .1427262
                     bos_away |   .0784663   .0033138    23.68   0.000     .0719713    .0849612
                     chc_away |   .0640756   .0032988    19.42   0.000     .0576099    .0705412
                  opening_day |   .4264648   .0063228    67.45   0.000     .4140721    .4388575
                 place_in_div |  -.0138547   .0005969   -23.21   0.000    -.0150246   -.0126848
                   games_back |  -.0030465   .0001841   -16.55   0.000    -.0034073   -.0026856
                     games_up |   .0029536   .0003217     9.18   0.000      .002323    .0035842
games_back_over_games_remaining |  -.0578722   .0051214   -11.30   0.000    -.0679101   -.0478344
                 doubleheader |   .0360508   .0054742     6.59   0.000     .0253214    .0467802
                 wins_last_10 |   .0030334    .000486     6.24   0.000     .0020809    .0039859
          runs_scored_last_10 |   .0000314   .0000681     0.46   0.644     -.000102    .0001649
                         rain |  -.0074325   .0048437    -1.53   0.125    -.0169261    .0020612
                      no_rain |   .0114145   .0030288     3.77   0.000     .0054781    .0173509
                  temperature |   .0010372   .0000777    13.34   0.000     .0008849    .0011896
                       streak |  -.0004069   .0002487    -1.64   0.102    -.0008944    .0000807
                        _cons |  -.0101009   .0314001    -0.32   0.748     -.071645    .0514432
--------------------------------+----------------------------------------------------------------
                       /sigma |   .1559423   .0004291                       .1551013    .1567833
-------------------------------------------------------------------------------------------------
  Obs. summary:          0  left-censored observations
                     66886     uncensored observations
                      2611 right-censored observations at perc_full>=1
```

5.3 Implications for Scheduling

Upon examination of the covariate coefficients in the Tobit regression, it is clear that the "advance" covariates dominate the regression. Suppose a business is scheduling staff for the night of a game two weeks in the future. The business guesses at the values of the significant "last minute" covariates in the Tobit regression: place_in_div, games_back, games_up, games_back_over_games_remaining, doubleheader, wins_last_10, no_rain, and temperature. Suppose the guesses for those covariates are off by 2, 7, 0, .2, 1, 4, 1, and 30, respectively. Suppose that all the errors are in the same direction, either all increasing the forecast of attendance or decreasing it. In that extreme case, the 2 week ahead forecast of attendance will only differ from the day-of-game forecast for attendance by 15% of stadium capacity. And recall that this is a very extreme case; it is extremely unlikely that the magnitude of the guesses will all be so large and will all be in the same direction. A 15% change in forecast amounts to 7500 fans in a 50,000 seat stadium. A business surrounding a stadium will not capture the business of every fan. If 1% of fans are captured for the business, this means a change in service load of 75 on the night on the game. This may be enough of a change to warrant an alteration of the initial staffing schedule. However, most nights, the change in forecast between 2 weeks ahead of the game and the day-of the game will only differ by 0-10 customers in the case of a business that attracts 1% of fans. Such a small change is unlikely to warrant a staffing alteration, as staffing schedule changes tend to decrease employee morale and lead to job burn-out.

With the "advance" covariates dominating the attendance forecast, businesses are encouraged to schedule staff weeks in advance. Alterations to the initial schedule may be made nearer to the game if the "last minute" covariates have changed drastically.

6 Conclusion

This paper has built and analyzed an econometric model of attendance at MLB games. The causes of game-to-game variation have been examined and quantified. Over 30 years of data were used to confirm that marquee matchups, winning home teams, weekend games during the summer, and special occasions (opening day, doubleheaders) help to bring more fans to the ballpark.

The model suggests that attendance predictors known long before the games starts (the date and time of the game, the opponent, past year performance of the home team) have a larger effect on attendance than predictors that only become clear near the day of the game (weather, recent performance of the home team). This conclusion is similar to that drawn by Beckman et al. (2011), which finds that such "last minute" covariates have been decreasing in importance over time.

Given the small effect of the weather and the recent team performance on attendance, businesses relying upon fan patronage are encouraged to schedule their staff in advance of game day as they normally would, using this prediction model as a guide for fan attendance. Best guesses as to the values of the weather and the recent team performance should be used in this advance forecast. As gameday nears, if the best guesses for the covariates turns out to be significantly wrong, it may be in the interest of the business to revisit its staffing plan and update based on recent events.

6.1 Future Work

This model was built for a course project in Fall 2014. It is missing a number of covariates that may improve performance. Future work to include these covariates may help the performance of the model or change the conclusions about the "advance" and "last minute" covariates: home and away starting pitchers, weather forecasts (in addition to realized weather), stadium closings (in addition to openings), rivalry matchups, holiday and holiday weekend indicators, and distance to the away team's stadium.

Additionally, a more developed staffing model needs to be created to examine the tradeoffs between setting the schedule early and waiting for more accurate covariate information. The anecdotal example given in Section 5.3 is a starting point, but needs to be formalized.

It would also be interesting to examine the effect of the covariates in this model both over time and across teams. The model could be fit on 5 year chunks of data at a time to examine the evolution of what is important to a generic baseball fan. Alternatively, all 30 years of home game data for a single team can be used to see how each city's fans differ in the importance they place on each regression covariate. Significant differences across fan bases would be important to note for businesses surrounding those teams.

References

ballparksofbaseball.com. 2014. Ballpark comparisons. URL `http://www.ballparksofbaseball.com/capacity.htm`. [Online; accessed 2-December-2014].

Baseball-Reference.com. 2013. 2013 major league baseball attendance & miscellaneous. URL `http://www.baseball-reference.com/leagues/MLB/2013-misc.shtml`. [Online; accessed 9-October-2014].

Beckman, Elise M, Wenqiang Cai, Rebecca M Esrock, Robert J Lemke. 2011. Explaining game-to-game ticket sales for major league baseball games over time. *Journal of Sports Economics* 1527002511410980.

Borland, Jeff, Jenny Lye. 1992. Attendance at australian rules football: a panel study. *Applied Economics* **24**(9) 1053–1058.

Butler, Michael R. 2002. Interleague play and baseball attendance. *Journal of Sports Economics* **3**(4) 320–334.

Donihue, Michael R, David W Findlay, Peter W Newberry. 2007. An analysis of attendance at major league baseball spring training games. *Journal of sports economics* **8**(1) 39–61.

Easton, Fred F. 2014. Service completion estimates for cross-trained workforce schedules under uncertain attendance and demand. *Production and Operations Management* **23**(4) 660–675.

Forrest, David, Robert Simmons. 2002. Outcome uncertainty and attendance demand in sport: the case of english soccer. *Journal of the Royal Statistical Society: Series D (The Statistician)* **51**(2) 229–241.

Fort, Rodney. 2004. Inelastic sports pricing. *Managerial and decision economics* **25**(2) 87–94.

García, Jaume, Plácido Rodríguez. 2002. The determinants of football match attendance revisited empirical evidence from the spanish football league. *Journal of Sports Economics* **3**(1) 18–38.

Hill, James Richard, Jeff Madura, Richard A Zuber. 1982. The short run demand for major league baseball. *Atlantic Economic Journal* **10**(2) 31–35.

Hur, Daesik, Vincent A Mabert, Kurt M Bretthauer. 2004. Real-time work schedule adjustment decisions: An investigation and evaluation. *Production and Operations Management* **13**(4) 322–339.

Kahane, Leo, Stephen Shmanske. 1997. Team roster turnover and attendance in major league baseball. *Applied Economics* **29**(4) 425–431.

Knowles, Glenn, Keith Sherony, Mike Haupert. 1992. The demand for major league baseball: A test of the uncertainty of outcome hypothesis. *The American Economist* 72–80.

Lemke, Robert J, Matthew Leonard, Kelebogile Tlhokwane. 2010. Estimating attendance at major league baseball games for the 2007 season. *Journal of Sports Economics* **11**(3) 316–348.

Madalozzo, Regina, Rodrigo Berber Villar. 2009. Brazilian football what brings fans to the game? *Journal of Sports Economics* **10**(6) 639–650.

McDonald, Mark, Daniel A Rascher. 2000. Does bat day make cents?: The effect of promotions on the demand for baseball. *Journal of Sport Management* **14**.

Meehan, James W, Randy A Nelson, Thomas V Richardson. 2007. Competitive balance and game attendance in major league baseball. *Journal of Sports Economics* **8**(6) 563–580.

Mehrotra, Vijay, Ozgür Ozlük, Robert Saltzman. 2010. Intelligent procedures for intra-day updating of call center agent schedules. *Production and Operations Management* **19**(3) 353–367.

Pinker, Edieal J, Robert A Shumsky. 2000. The efficiency-quality trade-off of cross-trained workers. *Manufacturing & Service Operations Management* **2**(1) 32–48.

Schmidt, Martin B, David J Berri. 2004. The impact of labor strikes on consumer demand: An application to professional sports. *American Economic Review* 344–357.

Thompson, Gary M. 1999. Labor scheduling, part 4 controlling workforce schedules in real time. *Cornell Hotel and Restaurant Administration Quarterly* **40**(3) 85–96.

Van Mieghem, Jan A. 2003. Commissioned paper: Capacity management, investment, and hedging: Review and recent developments. *Manufacturing & Service Operations Management* **5**(4) 269–302.

Welki, Andrew M, Thomas J Zlatoper. 1994. Us professional football: The demand for game-day attendance in 1991. *Managerial and Decision Economics* **15**(5) 489–495.

Welki, Andrew M, Thomas J Zlatoper. 1999. Us professional football game-day attendance. *Atlantic Economic Journal* **27**(3) 285–298.

Whitney, James D. 1988. Winning games versus winning championships: The economics of fan interest and team performance. *Economic Inquiry* **26**(4) 703–724.

Zhang, James J, Dale G Pease, Stanley C Hui, Thomas J Michaud. 1995. Variables affecting the spectator decision to attend nba games. *Sport Marketing Quarterly* **4** 29–40.