

Promotions and Opponents: Do They Matter to the M's?

Grant Camp 9922216
Econ 482
Autumn Quarter 2002

solaris@u.washington.edu

Executive Summary:

In the multi-million dollar industry of Major League Baseball attendance matters a great deal. Without fans, a team cannot survive. Fans are vital to both the revenue of the team, as well as the morale of the players. This paper builds a model for explaining the regular season attendance number for the Seattle Mariners.

The results of an ordinary least squares regression shows that although one would expect promotions and stronger opponents to increase attendance, this is not the case. Promotions do not seem to have any significant effect on attendance, nor does the opposing team's winning percentage.

Determining that promotions and opponent winning percentage does not affect attendance rates seems contradictory to common sense. However, winning percentage does not vary greatly across teams, which is most likely the cause for its insignificance. Also, promotions are often subsidized by outside advertising or are likely to be placed on games where lower than average attendance is expected. Since the Mariner's want a constant flow of income throughout the season, they could be using promotions to smooth their attendance numbers.

From the model we see that attendance can be explained based on the date alone. Weekend games will tend to have higher attendance rates than weekday games. Games later in the year will average greater attendance than games at the start of the season. And also each year since 2000 will have increased average attendance per game.

The Model

The purpose of this paper is to build a model to explain regular season home game attendance for the Seattle Mariners. Specifically, I will examine data from the 2000 season through the 2001 season. Variables that I believe affect attendance numbers are day of the week, month, opponent strength, and promotional give-a-ways. I expect that weekend games (no work to worry about), games later in the season¹, those against better opponents, and games with promotional give-a-ways will have higher attendance numbers.

The Data

Data collection was a fairly simple process due to the wonders of the Internet. All attendance numbers, dates, and opponent records were gathered off ESPN.com by way of “copy and paste” into an Excel document. The promotional give-a-way information was obtained through the Seattle Mariners website and through the Safeco Field information line. The data is limited to only three years due to two problems in data collecting. First, ESPN only publishes the past three years of data on their website, and second, Safeco Field seems to have difficulty operating a fax machine since I never received all of the information they agreed to send me. However, this is not really a problem since Safeco Field opened in 1999. Data before this time would have another variable, the field, and we all know the Kingdome was no beauty. So, in reality using only three years of data is not too bad.

¹ Although baseball “purists” claim that every game matters, I don’t think fans’ wallets agree. Later games seem to have more consequence than games played in April.

Other notes and caveats on the data and variables are as follows. Inter-league play was not included in the model because there were too few data points to make it into another variable. Post-season play was left out as well since the playoffs have an uncanny way of compelling fans to pay exorbitant amounts of money to see the Mariners lose (but that is another paper) causing every game to be sold-out, thus adding a variable of “post-season” would be uninteresting since we can empirically observe the consequence. As for the “date” variables, they are constructed in the following ways. DAY is a dummy variable that indicates weekend or not. For this model, a “weekend” game is defined as a Thursday, Friday, or Saturday game. Yes, the weekend starts on Thursday in many minds including mine for this paper. MONTH is just the month of the season. April being month 1 and September being month 6 (any October run-over games are considered September since they were supposed to be played then anyway). YEAR is pretty much self-explanatory. In the actual regression, I use (YEAR-2002) to get an idea of change in attendance over the past three years. OPR stands for opponent record in terms of winning percentage headed into the current game. The PROMO variable is another dummy variable denoting a give-a-way or promotion (take your kids to the ballpark day etc.). The final variable ATND is rather obvious since it is just the total attendance for that game. Take note that this measures the number of fans who made it through the door, not how many tickets sold for the particular game.

Estimated Model and Various Regressions

The model I began with was the following:

$$\mathbf{ATND} = a + \beta_1(\mathbf{DAY}) + \beta_2(\mathbf{MONTH}) + \beta_3(\mathbf{YEAR} - 2000) + \beta_4(\mathbf{OPR}) + \beta_5(\mathbf{PROMO})$$

Upon running the regression I obtained the following estimated coefficients²:

$$\text{ATND} = 39005 + 2066(\text{DAY}) + 1158(\text{MONTH}) + 2183(\text{YEAR} - 2000) - 367(\text{OPR}) + 312(\text{PROMO})$$

(23.09) (3.12) (6.28) (5.48) (-0.14) (0.47)

All of the coefficients have been rounded to whole numbers since talking about parts of people seems useless. The R^2 for this regression is 0.28. The estimated model says that weekend games will average about 2000 more fans than weekday games, each month the season progresses increases the average attendance by about 1100 and for the last three seasons attendance has increased about 2200 people per game. The relationships between attendance and each variable are about as expected except for OPR. The negative coefficient associated with OPR suggests that the better the opponents winning percentage is, the lower the attendance will be. This seems to contradict anecdotal evidence when the Yankees are in town. However, people do show up to boo Alex Rodriguez even though the Rangers' winning percentage is below .500. The only consolation comes from the fact that OPR appears statistically insignificant. From Figure 1 we see that all of the "date" variables are statistically significant with very high probabilities. In testing the hypotheses that the OPR and PROMO coefficients are equal to 0 (each one alone, not together), we find that neither variable is statistically significant. Since I attended a relatively random sample of games myself during the 2002 season I decided to re-run the regression only on the 2002 season. This of course means that the YEAR variable must be omitted since only one year is being examined. When running this regression some interesting results emerged³. It must also be noted that the R^2 drops to 0.11 in this regression so the model does not account for as much of the variation.

² Data in Figure 1 t-statistics listed below coefficients.

³ Data in Figure 2

However, during the 2002 season, OPR was most importantly positively correlated, and secondly, “more significant.” Instead of having an 88% probability of being 0, it only has a 41% probability. Why this occurs, I do not know. However, it does at least somewhat support my theory that the better an opponent the more people will turn out to see the games.

Conclusions

There are some irregularities in the data that do not agree with the expectations of the empirical model. It seems strange that the better an opponent is, the lower the attendance would be. Also, the insignificance of the PROMO variable seems to contradict common sense economics.

While my gut feeling would suggest that more people attend games against better teams the data does not seem to indicate this. There are several reasons that this could be the case. First and foremost is the correlation between winning percentage and opponents’ “strength.” While the numerical difference between .400 and .600 is not very great, it suggests a vast difference between the qualities of the two teams. This low variation between all teams⁴ is most likely the reason for the insignificance of the OPR variable. Also, 2001 was a bit of an anomaly for the Mariners. Since they had such a superior record to the rest of baseball, fans were attending games they wouldn’t have previously gone attended. Because of the added fervor due to the amazing season games against the Tampa Bay Devil Rays were probably selling out as easily as games against the New York Yankees. Possibly a better way to measure the effect of an opponent would be to create a dummy variable for each team and regress on those as well.

⁴ See Graph 1

However, this would approach the point where I don't have enough data points to support all of the right hand side variables.

An insignificant coefficient for the PROMO variable would suggest that any previous analysis done by the Mariners is incorrect. Economically speaking, one would assume that giving away posters and having discount packages should increase attendance thus expanding the fan base and increasing profits of the team. If this were not the case then there would be no incentive to do this. One explanation for this is that most of the give-a-ways are subsidized by companies advertising budgets and not by the Mariners. Another possibility is that give-a-ways are merely damage control. Giving items away could be a way to make people happier even if the Mariners lose the game, thus possibly encouraging them to come to future games. Most likely however, is that promotions are scheduled for games when low attendance is expected. If this were the case the promotions would cause less variation in the attendance numbers. Also, the promotions are relatively evenly distributed over the course of a season, and there is a maximum capacity to Safeco Field. Because of this, promotions are probably not going to have any effect near the end of the season when games are selling out based on the date alone. Perhaps promotions should be analyzed on a time basis as well.

There are certainly factors that are not accounted for such as weather, winning streaks, and rivalries. These variables are both difficult to collect and difficult to express quantitatively. Without these "extra" variables, I feel that this model does a decent job of explaining the regular season attendance for the Seattle Mariners.

Figure 1.

Dependent Variable: ATND
 Method: Least Squares
 Date: 11/13/02 Time: 15:19
 Sample: 1 216
 Included observations: 216

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	34637.91	1499.534	23.09911	0.0000
DAY	2065.900	662.0354	3.120528	0.0021
MONTH	1158.275	184.4888	6.278292	0.0000
YEAR-2000	2183.424	398.5743	5.478086	0.0000
OPR	-367.3483	2552.186	-0.143935	0.8857
PROMO	311.8534	657.9550	0.473974	0.6360

Figure 2.

Dependent Variable: ATND
 Method: Least Squares
 Date: 11/13/02 Time: 15:36
 Sample: 1 72
 Included observations: 72

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	39467.94	1607.971	24.54518	0.0000
DAY	2187.080	853.5451	2.562348	0.0127
MONTH	522.5228	233.8507	2.234429	0.0288
OPR	2328.733	2801.285	0.831309	0.4088
PROMO	41.70401		847.1227	0.049230

Graph 1.

