Do new stadiums increase attendance?

Ron Kamara
ECON 482
12/10/02
rkamara@u.washington.edu

Executive Summary

In recent years, many teams have lobbied for new stadiums on the basis that they increase attendance, and are thus vital to generating the type of income teams need to survive and remain competitive. The purpose of this study is to determine the actual effect of a new stadium on a team's per-game attendance, focusing on Major League Baseball.

This paper developed a model, using least squares regression analysis, for baseball attendance. The model included such factors as stadium capacity, attendance in the previous season, wins, excitement in the season (e.g. making the playoffs), ticket prices, and whether or not the team is playing in a new stadium.

According to this model, the two greatest factors in a team's attendance are the number of wins, and attendance figures from the prior season - suggesting a great deal of consistency in attendance figures. Excitement in a season also played a significant role.

Changes in the variables used in this analysis explained nearly 92% of the variation in per-game attendance figures. The model also performed remarkably well in a practical test, predicting the 2002 attendance figures with an average error of only 2,339.208 spectators per game.

This analysis estimated the effect of a new stadium on attendance to be as follows: in the first year of a new stadium, per game attendance increases by 10,226.34 spectators. This gradually dissipates, with increased attendance of 5402.727 in the second year of a new stadium, 3653.147 in the third, 2803.383 in the fourth, and 2472.083 in the fifth season.

While the model shows that there exists a clear and significant increase in attendance during the course of the first season in a new stadium, and that this will likely dissipate over subsequent seasons, there is still a great deal of statistical ambiguity regarding the actual rate of dissipation.

The Question

In recent years, many teams have pressed for new stadiums, arguing that new stadiums increase attendance and are thus vital to generating the type of income teams need to survive and remain competitive. The purpose of this study is to determine the actual effect of a new stadium on a team's per-game attendance. This paper focuses on Major League Baseball, because many of the new stadiums built in recent years were baseball stadiums, and also because the ballpark atmosphere is a bigger part of the experience in baseball than in other sports.

Data Collection

Attendance data, in the form of average per-game attendance, was collected from a sports almanac on www.infoplease.com as well as the 2002 Major League Baseball Report. The www.infoplease.com sports almanac was also used, along with www.espn.com, to collect the number of wins, playoff appearances, and other relevant performance-related statistics for each team. Statistics on stadiums, including capacity and when they were built, was collected from www.espn.com and individual teams' official websites. Ticket data was collected from Team Marketing Reports' Fan Cost Indexes, and is in the form of annual average ticket prices. Population data (which was entered with thousands as the unit of measure) was collected from the 2000 US Census.[1]

The Model

In order to answer this question, I used the following multiple regression using data for all MLB teams from 1997-2002[2]:

[1] US census data was available for the year 2000 only, and was used as a population estimate for the years 1997-2002. The 2001 Canadian Census was used as a population estimate for Canadian cities. For the purposes of this study, the population of Los Angeles was used for the Anaheim Angels, which play in a suburb of Los Angeles, and the population of Dallas was used for the Texas Rangers, which play in Dallas suburb Arlington. These population figures better reflect the teams' market.

[2] Data on the 1999 Seattle Mariners was excluded from this regression, because the Mariners opened a new stadium midway through the 1999 season.

ATT = B + B1*CAP + B2*EXCITE1 + B3*PATT + B4*WINS + B5*TIX + B6*TIX^2 + B7*STAD1 + B8*STAD2 + B9*STAD3 + B10*STAD4 + B11*STAD5 + B12*POP^2 + B13*YEAR_1997

The variable POP represents population, WINS the number of wins, TIX average ticket price, CAP stadium capacity, and YEAR_1997 the year-1997. EXCITE1 is a dummy variable representing additional excitement in a season, with 1 being assigned to teams that make the playoffs, as well as teams making runs at major records (e.g. the home run record) and expansion teams. STAD1 - 5 are all dummy variables, accounting for teams playing in the first year of a new stadium, the second year, and so on.

Results

Dependent Variable: ATT
Method: Least Squares
Date: 12/06/02 Time: 11:43
Sample(adjusted): 2 177
Included observations: 147
Excluded observations: 29 after adjusting endpoints

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-7008.377	3507.575	-1.998069	0.0477
CAP	0.021837	0.040307	0.541760	0.5889
EXCITE1	829.5344	851.5245	0.974176	0.3317
PATT	0.854092	0.041047	20.80752	0.0000
WINS	135.0150	30.55877	4.418208	0.0000
TIX	-66.66046	278.6015	-0.239268	0.8113
TIX^2	2.203456	6.196981	0.355569	0.7227
STAD1	10226.34	1273.977	8.027100	0.0000
STAD2	-3331.508	1003.514	-3.319840	0.0012
STAD3	-961.2787	1083.184	-0.887457	0.3764
STAD4	-316.7410	1123.383	-0.281953	0.7784
STAD5	77.73602	1086.562	0.071543	0.9431
POP^2	2.10E-05	1.72E-05	1.218280	0.2253
YEAR_1997	-307.5542	189.3006	-1.624687	0.1066

R-squared	0.917709	Mean dependent var	29203.91
Adjusted R-squared	0.909665	S.D. dependent var	9365.456
S.E. of regression	2814.854	Akaike info criterion	18.81360
Sum squared resid	1.05E+09	Schwarz criterion	19.09840
Log likelihood	-1368.800	F-statistic	114.0934
Durbin-Watson stat	1.939424	Prob(F-statistic)	0.000000

Before addressing the effect of new stadiums on attendance, it is necessary to explain the impact of the other variables. According to this model, the two greatest factors in a team's attendance are the attendance figures from the prior season and the number of wins. The strong impact of the number of wins is to be expected – when a team is successful, more people come to the games. The strength of the prior season's attendance demonstrates a great deal of consistency in attendance figures – this model estimates that a team will automatically retain over 85% of its attendance. This consistency is no doubt due in large part to consistency in team performance, but probably also reflects the impact of a team's fan base on attendance. For instance, I would infer that teams such as the Boston Red Sox and the New York Yankees, with their storied histories, have accumulated a large and consistent fan base and thus increased their attendance.

The level of excitement in a season, represented by EXCITE1, also has a strong impact. According to this model, a team that makes the playoffs or breaks a major record can expect roughly 830 extra fans per game. However, in reality this variable probably has little or no impact on the first half of the season, and a much larger impact towards the end of the season, when playoff contention and records become relevant.

The coefficients on TIX and TIX^2 reveal interesting results. Solving the first order condition for the partial derivative shows that attendance decreases as ticket prices increase so long as the average ticket price is less than $15.13, after which increases in ticket prices are associated with *increases* in attendance. However, the coefficients on TIX and TIX^2 were neither singularly nor jointly statistically significant.

Finally, this model estimates an increasing marginal return to population size. A team that plays in New York, the Major League city with the largest population, will draw roughly 1344.849 more fans per game than a team that plays in Tampa Bay – the smallest relevant city. However, it must be remembered that this is a comparison of the extremes and that the population coefficient was not statistically significant.

With regards to the validity of the model, I was encouraged by the high R^2 of .917709, which means that nearly 92% of variation in attendance is explained by changes in the model's righthand-side variables. However, I was still concerned about the model's accuracy due to the large number of variables that are not statistically significant. CAP, EXCITE1, TIX, TIX^2, STAD3, STAD4, STAD5, POP^2, and YEAR_1997 all are not statistically significant at the 5% level. I therefore tested the model by regressing on data through 2001 alone and using this regression to predict attendance values for all MLB teams in 2002. The model performed remarkably well, predicting the 2002 attendance figures with a minimum absolute error of only 128.223, a maximum absolute error of 7,144.358, and an average absolute error of only 2,339.208 spectators per game. Additionally, I tested the model for heteroskedasticity and serial correlation, and could not statistically reject that the model is homoskedastic and there is no serial correlation.

With the accuracy of the model confirmed, I was able to determine the additional attendance a team could expect from building a new stadium. The model predicts that per game attendance will increase by a whopping 10,226.34 spectators in the first year of a new stadium. The results for the next four years of a new stadium are a little harder to interpret. At first glance, it appears that the returns to attendance of a new stadium in years 2 through 4 are negative. It must be remembered, however, that the previous season's attendance is a factor in the model - with roughly 85% of the spectators returning automatically.

The correct interpretation is as follows: In the first year of a new stadium, an estimated 10,226.34 additional spectators attend each game. In the second year of a new stadium, 0.854092*10226.34 – 3331.508 = 5402.727 more people attend than would have had no new stadium been built. In the third season, 0.854092*5402.727 – 961.279 = 3653.147 more people attend each game than otherwise would have. Likewise, the estimated additional attendance is 2803.383 in the fourth year and 2472.083 in the fifth.

While the coefficients on STAD1 and STAD2 are statistically significant, the coefficients on the other three years are not. Therefore, I constructed 95% confidence intervals on the return to each year in a new stadium. For the first year this was easy to do, and the confidence interval is [7729.345, 12723.335]. This model therefore predicts, with 95% confidence, that per-game attendance will increase by anywhere from 7,729.345 to 12,723.335 spectators in the first year of a new stadium.

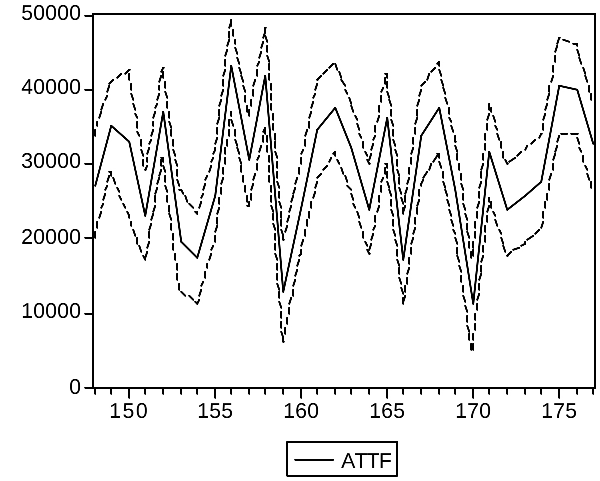
Confidence intervals for the other four years were more difficult to construct, because the return to attendance in those years involves a non-linear combination of coefficients. However, I was able to determine the standard error of each of these expected returns by running a Wald coefficient test in Eviews. Each of these non-linear combinations contains only one restriction; therefore, the generated F-statistic is the square of the t-statistic for the same combination. Thus, I was able to use the F-statistic and the expected value of the coefficient combination in order to compute the standard error. The 95% confidence intervals are [2214.646, 8590.808] for the second season, [-131.096, 7437.39] for the third, [-1317.772, 6924.538] for the fourth, and [-1898.458, 6842.624] for the fifth year.

It is clear from this model that there is a dramatic increase in per game attendance during the first season of a new stadium, and that this increase dissipates over time. However, the confidence intervals demonstrate that the rate at which the return of a new stadium to attendance dissipates isn't clear – to the point where it cannot be statistically ruled out that the attendance is *lower* than normal after the second season in a new stadium, although intuitively I don't believe that.

Thus, there is validity to franchise claims that the construction of new stadiums increases attendance at major league baseball games. However, although dramatic in the first season, this effect is in all likelihood limited to the short term and dissipates quickly over the next four seasons.

APPENDIX

Forecast:



Forecast: ATTF
Actual: ATT
Forecast sample: 148 177
Included observations: 30

Root Mean Squared Error	2922.165
Mean Absolute Error	2339.208
Mean Abs. Percent Error	9.990166
Theil Inequality Coefficient	0.048567
Bias Proportion	0.186210
Variance Proportion	0.072627
Covariance Proportion	0.741162

Coefficient tests:

For confidence interval construction:

Wald Test:
Equation: Untitled

Null Hypothesis:	C(4)*C(8)+C(9)=0		
F-statistic	11.03264	Probability	0.001156
Chi-square	11.03264	Probability	0.000895

Wald Test:
Equation: Untitled

Null Hypothesis:	C(8)*(C(4)^2)+C(9)*C(4)+C(10)=0		
F-statistic	3.580047	Probability	0.060651
Chi-square	3.580047	Probability	0.058478

Wald Test:
Equation: Untitled

Null Hypothesis:	C(8)*(C(4)^3)+C(9)*(C(4)^2)+C(10)*C(4)+C(11)		
F-statistic	1.777620	Probability	0.184721
Chi-square	1.777620	Probability	0.182442

Wald Test:
Equation: Untitled

Null Hypothesis:	C(8)*C(4)^4+C(9)*C(4)^3+C(10)*C(4)^2+C(11)*C(4)+C(12)=0		
F-statistic	1.229045	Probability	0.269594

Chi-square	1.229045	Probability	0.267593

For joint significance on TIX and TIX^2 coefficients:

Wald Test:
Equation: Untitled

Null Hypothesis:	C(6)=0		
	C(7)=0		
F-statistic	0.137321	Probability	0.871814
Chi-square	0.274642	Probability	0.871690

Serial correlation test:

Breusch-Godfrey Serial Correlation LM Test:

F-statistic	0.001617	Probability	0.967983
Obs*R-squared	0.001801	Probability	0.966150

Heteroskedasticity test:

White Heteroskedasticity Test:

F-statistic	0.644294	Probability	0.969699
Obs*R-squared	65.46420	Probability	0.895341