David J. Berri, Martin B. Schmidt and Stacey L. Brook
Stars at the Gate: The Impact of Star Power on NBA Gate Revenues

David J. Berri, Martin B. Schmidt and Stacey L. Brook

Jack Dolitsky

The NBA has not been able to replicate the competitive balance of other major sports leagues. The paper that I am replicating does not look at what leads to this level of competitive balance, rather the effect it has on consumer demand. The question, as stated by the paper is, "How do these NBA teams still maintain demand in the face of the certainty of an unwelcomed outcome." In other words, the question is how the NBA can keep consumers wanting their product, even if the consumer knows the what the end result of the game will be. This question is important because it can be a major factor in business decisions and rule changes. If it is found that there is a negative effect on demand caused by disparity, the NBA can implement free agency rules, salary cap rules etc. in order to curb that effect.

The main findings of the paper look at what attracts consumer demands. It was suggested in a previous paper that was looked at that by Hausman and Leonard (1997) that star power was the largest factor in consumer demand. In this paper, star power was shown to be significant, but it was concluded that winning overall has a larger effect on gate revenues. One of the questions that the authors were interested in researching further was if the true factor in consumer demand the opponent team's star player is rather than the home team's star player.

I added two variables to my replication. I added one variable accounting for if a team was the runner up for winning the championship, and one variable for the highest individual BPM value for each team, filtered for players who have played over 1000 minutes. The reason I added the first variable is because there was already a variable accounting for championships won in the last 20 years, but only 8 different teams have finals appearances in that timeframe, and teams that were the runner up also had a lot of playoff success and should be accounted for. I also added the BPM (box plus-minus) variable because the paper looks at all-star votes to account for consumer demand when it comes to star players, but I wanted to see if there was an effect based on how good (based on the given player's BPM statistic) the team's star player actually performed on the court.

The paper I was replicating did not have the data for gate revenues available anywhere online for the 1993-1996 seasons, so I found later data from the 2011-2014 season so I could get as close to getting all of the variables as possible. I got the gate revenues from Statista. I got variables for all the team wins for the regular season and playoffs from basketball-reference. I also created my own variable from that data to calculate the competitive balance, which is a variable I created from the basketball reference data. I also scraped the aforementioned variables that I added from basketball reference. Another variable that I scraped from basketball references were all-star votes. Since the all-star votes data frame only had the player and not the team, I had to match up the players with the roster data I scraped to get the BPM variable. I used the roster data to create the roster stability variable. The roster stability variable is the percentage of minutes played of players on a given team in a given season, over the same percentage of those same returning players in the following season. The attendance numbers were scraped from ESPN. That included the variables for the stadium capacity (which I created from average attendance and percent at capacity), and total attendance numbers. There was not enough that explained the variable for percent of teams at capacity so I could not get that variable. I got the population data from the Census Bureau, which was the population for the team's city for that season. I got the per capita income data from the deptofnumbers.com. I also could not get the white ratio variable because there was no dataset for it online, and I would have had to manually find every player on every roster that fit the criteria for the variable.

$$\mathbf{Y}jtv = log(\prod \mathbf{X}jtv) + \varepsilon + Xfe$$

The above equation is the main model used in the paper. It is a double log model, in which it takes the gate revenues of a team in a given league in a given period. It is the log of the product of the independent variables. It adds the error term and the fixed effects for the year. This means if you change the independent variable by a certain percent the independent variable will be expected to change by $\beta 1$ percent. One of the biggest limitations of the paper is omitted variable bias (although my replication does not include the same years as the authors, it would be likely that there is same effect). When I added my created variables, I noticed that some variables changed in terms of coefficient, as well as statistical significance, including competitive balance, which I consider one of the most important variables. Also, the paper doesn't discuss how it handles the double log model, and since some of the variables have values of zero, I had some trouble figuring out how it was handled. My solution was to add 1 to each variable that had a 0 in it, so the values were changed very slightly but it was able to work in the model.

As I mentioned previously, I added variables for championship runner ups and for the player who had the top BPM value on the team's BPM. Since the original paper had dummy variables for some superstar players such as Michael Jordan, Charles Barkley, and Shaquille O'Neal, it would be helpful to add a variable for star players. For the runner up variable, I handled it similarly to the championship variable. If a team won a championship in the prior season, they were assigned a value of 20, 19 if it was two years prior, etc. I used the variable in the linear regression model, the double logged model, and the table for marginal revenue.

The following regression results are the double logged model, the linear model, and the marginal revenue table respectively.

	Dependent variable:
	log(Gate Revenues)
log(Wins)	0.296*
	(0.160)
log(Playoff Wins)	0.052
	(0.053)
log(Playoff Wins)	0.133***
	(0.046)
log(Championships)	0.112***
	(0.028)
log(All Star Votes)	0.009
	(0.007)
log(Stadium Capacity)	0.984
	(0.687)
log(Roster Stability)	-0.274
	(0.171)
log(Population)	0.030
	(0.055)
log(Income)	0.618***
	(0.216)
log(Competitive Balance)	-0.528*
	(0.303)
2012	-0.201*
	(0.116)
2013	0.038
	(0.095)
2014	0.066

	(0.095)
Constant	0.312
	(6.556)
Observations	116
Log Likelihood	-37.630
Akaike Inf. Crit.	103.260
Note:	*p<0.1; **p<0.05; ***p<0.01

	Dependent variable:
	Gate Revenues
Wins	31,242.210
	(180,291.000)
Playoff Wins	739,920.600*
	(418,149.900)
Playoff Wins Lagged	459,468.000
	(392,556.500)
Championships	314,214.600***
	(97,963.370)
All Star Votes	6.746***
	(1.705)
Stadium Capacity	3,152.729**
	(1,301.265)

Roster Stability	-6,878,926.000
	(11,078,948.000)
Population	4.424***
	(1.211)
Income	134.467
	(243.566)
Competitive Balance	-2,559,089.000
	(1,571,326.000)
2012	-4,699,683.000
	(4,301,949.000)
2013	7,135,134.000*
	(3,696,932.000)
2014	8,065,584.000**
	(3,678,939.000)
Constant	-26,115,801.000
	(24,616,088.000)
Observations	116
R^2	0.711
Adjusted R^2	0.674
Residual Std. Error	13,125,791.000 (df = 102)
F Statistic	19.326*** (df = 13; 102)
Note:	*p<0.1; **p<0.05; ***p<0.01

Coeffiecient or Elasticity	P-value	Mean	$ Average Marginal Revenue
log(Playoff Wins)	0.04879173	3.766667	162683.4139
log(Runner Ups)	0.00002392303	7.116667	117083.2522
log(Competitive Balance)	0.007937217	7.588107	-725402.6334
log(Championships)	0.004021348	8	70636.135
log(Wins)	0.04161753	38.991667	70185.1686
log(Incone)	0.004193179	36271.18966	114.2071

Since my data is very different from the authors due to different years, I will focus on the double log model and describe what is drastically different, and what is similar. One of the most glaring differences is the significance of the all-star votes variable. While the authors variable was statistically significant at the 1-percent level, my variable was not statistically significant at all, which suggests that in the 90s, consumer demand was based on fan-favorite players more than it was in the early 2010s. This was the same case for population. In terms of similarities both wins, and championships were statistically significant for both of our data sets with positive coefficients. This would be expected as team success is discussed in the paper as being a major driving force in consumer demand. The following table will be the logged model with my added variables.

	Dependent variable:
	log(Gate Revenue)
log(Wins)	0.386**
	(0.187)
log(Playoff Wins)	0.069
	(0.049)

log(Lagged Playoff Wins)	0.086**
	(0.043)
log(Championships)	0.080***
	(0.027)
log(Runner Ups)	0.117***
	(0.026)
log(All Star Votes)	0.011*
	(0.006)
log(Top BPM)	-0.059
	(0.086)
log(Stadium Capacity)	0.477
	(0.644)
log(Roster Stability)	-0.296*
	(0.158)
log(Population)	0.053
	(0.052)
log(Income)	0.584***
	(0.199)
log(Competitive Balance)	-0.776***
	(0.286)
2012	-0.212*
	(0.111)
2013	0.025
	(0.089)
2014	0.049
	(0.088)

Constant	5.517
	(6.162)
Observations	116
Log Likelihood	-27.131
Akaike Inf. Crit.	86.263
Note:	*p<0.1; **p<0.05; ***p<0.01

Something that is very interesting is that competitive balance becomes more significant which suggests it has a correlation with my added variables. As discussed before, this likely demonstrates omitted variable bias in the original paper. My variable for runner ups is even more significant than champions, with a larger coefficient. This could mean that since both the champions and runner ups are successful, there is not a large enough discrepancy in success for championships to have a significant positive effect on consumer demand more than runner up. The top BPM variable was not significant, which is almost consistent with what was said in the paper. The paper claims that stars have a small effect but not as much as team success, but my results show they have little to no effect, especially with the coefficient being negative.

References

2021. [online] Available at: <https://www.statista.com/statistics/> [Accessed 12 April 2021].

Basketball-Reference.com. 2021. *Basketball Statistics and History | Basketball-Reference.com*. [online] Available at: <https://www.basketball-reference.com/> [Accessed 12 April 2021].

Berri, D., Schmidt, M. and Brook, S., 2004. *Stars at the Gate: The Impact of Star Power on NBA Gate Revenues*. [online] Available at: <https://www.researchgate.net/publication/247739515_Stars_at_the_Gate_The_Impact_of_Star_Power_on_NBA_Gate_Revenues> [Accessed 12 April 2021].

Deptofnumbers.com. 2021. *Chicago Illinois Household Income | Department of Numbers*. [online] Available at: <https://www.deptofnumbers.com/income/illinois/chicago/> [Accessed 12 April 2021].

ESPN.com. 2021. *2011-2012 NBA Attendance - National Basketball Association - ESPN*. [online] Available at: <http://www.espn.com/nba/attendance/_/year/2012> [Accessed 12 April 2021].