Alex Scarcelli
Professor Van Kammen
Econ 325
04/28/2021

Project Write Up

Most basketball fans would rather watch a high scoring, fast paced game than a defensive, slow paced game. Early on in the NBA's history, games were played with a focus on getting the ball in the paint for easier shots and layups. However, after the introduction of the three-point shot, the game style changed. In the modern-day NBA, games are more fast paced and high scoring, with the three-point line becoming a major weapon teams use to win games.[2] Now, we are seeing three-point oriented players, such as Steph Curry and Klay Thompson, emerging as superstars. As more points are scored in the modern NBA games, the question becomes if fans enjoy these higher scoring games more than low scoring and defensive games? So, this paper attempts to ask the question, "does scoring more points per game increase fan attendance?"[2]

In Todd Copenhaver's, "Does NBA Attendance Respond to Increased Offensive Quality and Production," he predicts that if the NBA knew the fans preferences, their recent changes towards offensive production should have resulted in increased attendance. He predicts fans like more high scoring and fast paced offensive games. To evaluate this hypothesis, he uses three measures of offensive productivity: fast break points, offensive efficiencies, and turnovers per game. Using data from the 2007-2008 NBA season, he finds significant results that turnovers, fast break points, offensive efficiency, and points scored for the home team are all positive determinants of attendance. However, he also found evidence that away team points scored has a significant effect on attendance, but game speed does not have a significant effect. Copenhaver reached these conclusions by first controlling for the home team's previous season performance. He then fixes multiple variables of the home team, including locational qualities and ticket prices. He also adds the fixed effects of the away team to his equation. He then controls for multiple variables that could affect the attendance. These variables include weekend and weekday games, points in home games, points in away games, if Lebron, Kobe, or Kevin Garnett are playing, offensive efficiency at home and away, and finally the win percentage at home and away. In all three of his regressions, his fixed effects were statistically significant. Of all the star players, Kevin Garnett had the biggest impact on game attendance, drawing in 2900 more fans per game. Weekend games brought in 1200 more fans than games played in the week. All the measures of offensive quality were significant for the home team, but only a few were significant for the away team.

For my analysis, I compared different variables than Copenhaver to get their effects on average attendance. Instead of using the average points per game of both the home and away teams, I just used the average points per game of each team for that season. I controlled for the variables of team win percentage that season, the city population size that year, and the average ticket price that season. In one of my regressions, I fixed for the team's previous season success, and in another I added a lag for the previous season's win percentage. While Copenhaver used data from the 2007-2008 NBA season, I used more recent data from the seasons ending in 2017-2019. A noticeable difference in the data is the increased mean PPG in my data compared to Copenhaver's. This can be explained by the increasing number of three pointers attempted in the modern NBA. The summary of my data compared to Copenhaver's is shown below.

My Data Summary:

Variable	Obs	Mean	Std. Dev.	Min	Max
AVG	90	17910.96	926.7943	16088	19856
PPG	90	107.8611	4.801467	97.9	118.1
winpct	90	.4999778	.1425535	.207	.817
citypop	90	1234207	1045798	194188	3967000
avgticket	90	65.51667	37.1895	26	219

Copenhaver's Data Summary[1]:

Table 1

Variable	Mean	Std. Dev.	Min	Max
attendance	16914.04	3213.44	8393.00	22778.00
psptsaway	98.74	3.97	93.70	110.20
pswinsaway	40.92	10.62	22.00	67.00
KG	0.04	0.19	0.00	1.00
lebron	0.03	0.18	0.00	1.00
kobe	0.04	0.19	0.00	1.00
weekend	0.47	0.50	0.00	1.00
winpaway	0.50	0.17	0.10	0.92
ptsaway	99.03	5.10	85.44	111.10
fast_breakaway	12.19	3.29	5.44	22.36
turnoversaway	14.09	1.31	10.88	18.40
oeffaway	108.50	4.24	92.46	116.50
winphome	0.45	0.15	0.09	0.80
ptshome	97.93	4.73	85.36	111.06
fast_breakhome	12.12	3.37	5.88	22.84
turnovershome	14.24	1.26	11.10	18.08
oeffhome	107.24	3.88	93.24	115.62

For my regressions, I mostly got similar results to Copenhaver. However, the negative coefficient for the average PPG was a major difference. From my regression, I determined that for one more point per game, average attendance for that team will decrease by around 31 people with a standard deviation of around 12 people. For a 1% increase in win percentage, the average attendance is expected to increase by around 2948 people per game. An extra person in the population would increase average attendance barely, by 0.00017 of a person. An increase of the average ticket price by $1 would increase average attendance by around 7 people. The effect of average attendance per game could be biased negatively if the team had a successful season last year. For this reason, I control the home teams previous season success. I also add a lag of winning percentage in my third regression to determine if we would see the results of the previous season's success in the average attendance of the current season. My results and Copenhaver's results are shown below.

Regular Regression:

AVG	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
PPG	-31.49742	11.95838	-2.63	0.008	-54.93542	-8.059423
winpct	2948.134	462.228	6.38	0.000	2042.184	3854.084
citypop	.0001712	.0001293	1.32	0.185	-.0000822	.0004245
avgticket	7.050015	2.486049	2.84	0.005	2.177448	11.92258
_cons	19161.12	1189.709	16.11	0.000	16829.34	21492.91
sigma_u	695.55249					
sigma_e	325.65631					
rho	.82020365	(fraction of variance due to u_i)				

Fixed Season Regression:

AVG	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
PPG	-29.62012	17.24248	-1.72	0.086	-63.41475	4.174511
winpct	2929.437	528.5597	5.54	0.000	1893.479	3965.395
citypop	.000172	.00013	1.32	0.186	-.0000828	.0004268
avgticket	6.769135	2.629543	2.57	0.010	1.615325	11.92294
YEAR						
2018	70.65544	88.97249	0.79	0.427	-103.7274	245.0383
2019	15.03743	131.596	0.11	0.909	-242.886	272.9608
_cons	18956.83	1687.448	11.23	0.000	15649.49	22264.17
sigma_u	695.08262					
sigma_e	328.65222					
rho	.81728514	(fraction of variance due to u_i)				

Lag in Win Percentage Regression:

AVG	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
PPG	-30.51671	12.55513	-2.43	0.015	-55.12431	-5.909116
winpct						
--.	2341.404	550.9861	4.25	0.000	1261.491	3421.317
L1.	328.3926	505.0557	0.65	0.516	-661.4983	1318.283
citypop	.0001249	.0001336	0.93	0.350	-.0001371	.0003868
avgticket	6.734463	2.709222	2.49	0.013	1.424486	12.04444
_cons	19287.87	1291.699	14.93	0.000	16756.19	21819.56
sigma_u	717.90194					
sigma_e	263.21567					
rho	.88150068	(fraction of variance due to u_i)				

Copenhaver's Regression[1]:

Table 3

	Regression 1			Regression 2			Regression 3		
Variables	Coef.	t-stat	Elasticity	Coef.	t-stat	Elasticity	Coef.	t-stat	Elasticity
winpaway	-374.85	-0.63	-0.011	-599.67	-1.10	-0.018	-451.47	-0.67	-0.013
ptsaway	**41.00**	2.16	0.240	28.12	1.35	0.165	-	-	-
turnoversay	34.75	0.63	0.029	-	-	-	-	-	-
fast_breakaway	-	-	-	**50.40**	2.37	0.036	-	-	-
oeffaway	-	-	-	-	-	-	32.01	1.28	0.205
psptsaway	-7.60	-0.37	-0.044	-18.06	-0.87	-0.105	19.64	1.24	0.115
pswinsaway	**44.05**	5.94	0.107	**45.96**	6.21	0.111	**38.70**	5.54	0.094
winphome	**2267.00**	1.89	0.060	**2932.13**	2.43	0.078	2001.02	1.63	0.053
ptshome	**187.25**	4.65	1.084	**213.86**	4.55	1.238	-	-	-
turnovershome	**-625.34**	-6.21	-0.527	-	-	-	-	-	-
fast_breakhome	-	-	-	**168.49**	2.15	0.121	-	-	-
oeffhome	-	-	-	-	-	-	**287.98**	7.93	1.826
KG	**2888.18**	7.54	0.006	**3152.99**	8.14	0.007	**2880.31**	7.66	0.006
lebron	**814.84**	2.75	0.002	**826.86**	2.75	0.002	**903.71**	3.00	0.002
kobe	**1613.13**	5.39	0.003	**1829.70**	6.01	0.004	**1703.72**	5.74	0.004
weekend	**1204.22**	11.57	0.033	**1167.34**	11.03	0.032	**1206.27**	11.43	0.034
Constant	291.18	0.06	-	**-11345.05**	-2.62	-	**-22402.11**	-5.50	-
Fixed Effects	F(23, 828)	84.02		F(23. 830)	96.49		F(23, 828)	103.09	
F-test		88.03			84.37			90.02	
R^2		0.788			0.781			0.78	
Adj R^2		0.779			0.773			0.77	

Observations: 864

Although I had some similar aspects of my replication of Todd Copenhaver's estimates, I had had some very different results. While Copenhaver used data only from the 2007-2008 season, I used data over three seasons, 2016-2017, 2017-2018, 2018-2019. Although we had different seasons, our unit of observation, the NBA, was the same. The teams were the same except for a few that changed names. The big difference between our data sets was that since he only used one season, he only had a size of 30 teams for his sample. I used three different seasons, which gave me a larger sample of 90, or 30 teams for each season observed.

For our regressions, we both had positive coefficients for win percentage. His win percentage coefficient (2267) and mine (2948) were very similar with similar magnitudes. This positive coefficient is easily explained, as more fans will go to see their team play if the team is winning a lot of games. Fans want to see good teams play. However, something not so easy to explain is the difference between Copenhaver's PPG coefficient (187.25) and mine (-31.4972). His estimate of PPG is both qualitatively and quantitively different than mine. An explanation for his coefficient is also simple. Theoretically better teams score more points which increases the attendance or fans like high scoring, offensive games which also increase the attendance. My estimate is harder to explain. At first, I thought this was a fluke in my regression. I tried throwing different variables and controls to get a positive PPG coefficient. I fixed and put a lag on previous season success and got the same result. I also tried controlling for weekend and weekday games and still got a negative coefficient. After thinking about it for a while, I came up with a few theories for this negative coefficient. Firstly, fans are more interested in seeing low scoring games. Theoretically, low-scoring games tend to be closer with better defense. Therefore, fans could prefer attending closer games. Secondly, fans also might prefer to see games with better defensive play. Better defense means less scoring, which would explain why fans would want to see more low scoring games. Ultimately, my intuition about fans' preferences was wrong. I originally predicted that fans would prefer faster paced, high scoring games. However, it turns out they actually don't like higher scoring games. But why was Copenhaver's coefficient positive and mine negative? The difference could be due to fans changing their preferences over time. In the 2007-2008 season, fans obviously preferred attending these higher scoring games. However, in the modern NBA fans preferences have changed. They either got sick of blow outs and started preferring closer games, or they started preferring defensive games over offensive games. Some other reasons for the differences are the different variables we controlled for. He controlled for more aspects of the game, such as star players and turnovers, while I controlled for Economic variables such as city population and average ticket price. At the end of the day, these estimates only analyze a small portion of what impacts a game's attendance. More theories and variables need to be tested in order to better understand the factors that affect these games.

Works Cited

[1]Copenhaver, Todd, "Does NBA Attendance Respond to Increased Offensive Quality and Production?" (2009). *Economics Honors Projects*. 21. https://digitalcommons.macalester.edu/economics_honors_projects/21

[2]Scarcelli, Alex, "Topic Proposal" (2021). *Purdue University Econ 325 Project Topic Proposal*

StataCorp. 2019. *Stata Statistical Software: Release 16*. College Station, TX: StataCorp LLC. (Used for regressions)