Gambling Your Chances at Future Success in the NBA: An Econometric Approach to Tanking in the Playoff Race

Nathan Osborn

Alex Johnson

José Melián Theís

Econ 488

Professor Denning

I. Introduction

In the year 1985, the National Basketball Association (NBA) created and instituted the draft lottery. In this system, the teams that do not make the playoffs (which consist of the top 8 teams from the Eastern and Western Conferences) are entered into a lottery to determine the order in which the teams will be able to draft new players entering the league. The lottery odds are decided according to teams' records: the worse it is, the better their probability at securing a top pick in the draft. The teams that have a top pick then have a better chance at selecting the next superstar, thus improving their future results. Due to this format, general managers of teams that have little to no chance of making the playoffs in a given year will deliberately make their team worse in efforts to increase their chances of having a top pick in the draft, a process known as tanking. Much has been said about the prevalence of tanking. In prior research, Yigal Gerchack et al discuss tanking in their paper, "The Evolution of Draft Lotteries in Professional Sports: Back to Moral Hazard?" while Brian P. Soebbing et al discuss it in their paper, "Do Gamblers Think that Teams Tank? Evidence From the NBA." Both of these papers find evidence that tanking most likely exists, but they only focus on the actions of teams at the bottom of the standings. However, teams other than those at the bottom of the standings may also decide to tank for reasons that are explained in what follows.

In the NBA, playoff matchups are determined by conference rank and the disparity in skill level between the 1 and 2 seeds and the 7 and 8 seeds is quite large, with the higher seeds winning 93% of the time (Cohen). If a team's goal is to win the championship, qualifying for the playoffs as a 7 or 8 seed makes little difference in their odds of winning. The lottery, however, provides the opportunity of finding a player that will improve their team's performance in the future. Thus, teams may be better served to miss the playoffs on purpose (tank) and enter the draft lottery. A counter argument suggests that the experience of playing in the playoffs is beneficial to teams, so teams that make the playoffs, even as a 7 or 8 seed, will experience greater performance in the future. This paper will attempt to show the effect that drafting in the lottery has on future winning percentage for teams on the edge of making the playoffs.

II. Data

For this paper we gathered data from a combination of several different online data repositories, including landofbasketball.com, kaggle.com, and data.world. Multiple datasets that were relatively easy to format and clean came from both kaggle.com and data.world. However, to accurately gather the data from landofbasketball.com, we created a web-scraping program to concatenate multiple tables from the site into a single dataset. Being left with multiple incongruent datasets, we then used Python to create code to further clean and organize the data, before uploading the data to Stata to run the various pertinent regressions. Using its Pandas package, the implementation of Python code better facilitated the creation of the one contiguous dataset than other methods considered.

We have collected data from the 1984 to 2015 NBA seasons, since 1985 was the first year that the NBA instituted the lottery (each NBA season spans two years, in our paper we will refer to the season using only the year in which the season began). Firstly, the outcome of interest will be the teams' winning percentage in the year following this observation year, as is consistent with the guiding research question. The first key explanatory variable of interest will be each team's conference rank. As stated previously, playoff qualification is determined by conference rank, with the top 8 teams qualifying for the playoffs, where the rank is established by a given teams' winning percentage over the course of the 82-game season. This variable will always be within the range of 1 and 15 (though the range may be smaller than this), with 1 meaning the team has the best winning percentage in their conference, and 15 meaning the team has the worst (or is tied for the worst, at the very least). The next variable is the team's draft position. This variable will be between 1 and 30, with 1 meaning the team has the first pick in the first round of the draft and 30 meaning they have the last pick in the first round. Though there are two rounds in the NBA Draft, we focus only on the first as second round players are not guaranteed a spot on the roster. Now, since teams may voluntarily trade future draft picks for players, we will use the draft position data to create a dummy variable of whether a team picks in the lottery: 1 means they have a lottery pick, 0 means they do not, all to note whether a team that missed the playoffs actually gets a pick in the lottery. Next, we add a control variable of total salary given to players of each team during each NBA season of interest, as higher spending teams generally perform better due to spending on better players. In addition, we add a control for teams' roster continuity that calculates the percentage of minutes played in a season by players that were in the roster the previous season. This variable can be a good control given that players that have competed together for a long time have more chemistry. Lastly, we add a team fixed effects to control for those aspects of teams that don't change from season to season. A table of summary statistics is given in Table 1.[1]

These being the variables used, we believe this data to be ideal for the purposes of this analysis since it provides the ideal conditions for the preferred identification strategy as well as being directly related to the research question. We note however, that the data outlined likely does not account for all variation between a given team from season to season. Our model would likely benefit from added controls that account for this potential variation, similar to the other control variables mentioned earlier.

III. Identification Strategy

To estimate the effect the lottery has on future winning percentage we will use a fuzzy regression discontinuity (RD). The outcome variable is the winning percentage in the next year. The treatment variable will be the dummy variable for being in the lottery. The running variable will be conference rank. This will be a fuzzy RD because missing the playoffs does not guarantee that a team picks in the lottery, because of trades. Thus, teams that trade away their draft pick in the lottery, but miss the playoffs represent never takers, teams that make the playoffs but receive draft picks from the lottery represent always takers, and teams that possess their own draft pick and miss the playoffs are the compliers, which are the teams our study will focus on.

The first assumptions needed for this analysis are those of the regression discontinuity: balance of covariates and continuity of density. The first was calculated for our covariates: salary and team continuity. When we regressed each variable [1]against the main estimating function, there were not any significant changes across the threshold, suggesting a strong balance of covariates. This is shown in Table 2 of the appendix. We then tested for a discontinuity in density using the variable of games behind the eight seed. This however shows that there is a discontinuity, as seen in Figure 1 of the appendix. To control for this bunching, we use

[1] All referred-to tables and figures can be found in the Appendix section.

conference rank as our running variable rather than games behind the 8 seed. In this case there will never be a discontinuity in density since the number of teams across each side of the threshold is the same.

Next, we discuss the four assumptions for the instrumental variable analysis: first stage, independence, exclusion restriction, and monotonicity. To start off, the first stage represents the correlation between the position at the end of the season and the lottery pick. The resulting calculation was moderately strong, which leads us to conclude that the first stage is likely strong enough for the purposes of our analysis. For robustness, we show the first stage using draft pick position rather than an indicator for being in the lottery. This has a similarly significant effect, as shown graphically in Figures 2a and 2b, and numerically in Table 2.

Second, the independence assumption must show that being above or below the threshold for making the playoffs is as good as random. First, teams can only manipulate the games they participate in. Using the threshold of making the playoffs makes the independence assumption hold because teams are dependent on other teams as well if they are to win. Even though a team can manipulate individual games, the given team still depends on what other teams do, thus making playoff entry as good as random, for our purposes. Further, it can be reasonably believed that teams near the threshold will not purposefully tank and instead try to make the playoffs. We assume that this is the case since those teams would then be missing out on bonuses and other financial incentives, as well as giving these players important opportunities for experience (for the 2019-2020 season, playoff teams split a combined $23,287,266, for example) (Reynolds).

Third, the exclusion restriction assumption would state that missing the threshold of making the playoffs only affects future success the next year through drafting in the lottery. We assume this is true as if a team misses the playoffs, the only consequence is receiving a lottery pick, they do not play any more games that season and there is no free agent recompense or anything similar that may affect the next year's winning percentage. Thus, only through the lottery pick can missing the playoffs affect the next year's winning percentage.

Lastly, monotonicity assumes that if a team qualifies for the playoffs, there is not a higher chance to get a lottery pick. This assumption should generally hold due to the draft lottery system itself. The teams that qualify for the playoffs never have their own pick in the lottery, and we assume with a high degree of certainty these teams aren't trading frequently for lottery picks, since only through trades can teams receive a higher pick, which we control for.

Our estimating equation will be:

$$futurewin = \beta 0 + \beta 1 lottery + \beta 2 lottery * rank + \beta x + \lambda + \varepsilon$$

Where *futurewin* is the outcome variable of winning percentage in the next year. *Lottery* is a dummy variable for if they are in the lottery or not which is the treatment. *Rank* is the conference rank which is the running variable. *Lottery*rank* represents the interaction of *lottery* and *rank* which is the flexible function. *X* is the vector of covariates from the list of controls that were discussed previously. λ will represent team fixed effects, with assumptions involving these fixed effects being explained in the following paragraph. Finally, ε is the pertinent error term for the regression model.

In introducing team fixed effects to the model, we are making an additional assumption that there were no contemporaneous shocks during any year that affected some teams but did not affect others. There are several possible events that we consider in checking for contemporaneous shocks, such as the 1998 and 2011 NBA lockouts that significantly reduced the number of games played by all teams in the given seasons. Though the lockouts themselves likely do not constitute contemporaneous shocks (since they affect all teams equally), there might be scheduling concerns that give some teams a harder schedule, for example. However, we assume preliminarily that none of the potential shocks constitute a violation to the team fixed effects utilized in the model, since none of the events considered seemed to greatly influence teams unequally.

IV. Results

In general, our analysis finds that there are no statistically significant results affecting future winning percentage through the instrumented regression discontinuity design. It should be noted, however, that these results, in certain places, do approach levels of significance, and perhaps, with a larger dataset, they might be significant. In the context of our analysis, we will discuss specifically what these key findings were, including an explanation of their interpretation and potential ramifications. Next, we perform several robustness checks, first where we observe the fuzzy regression discontinuity at differing bandwidth levels, and then by performing a separate regression which is simply the reduced form estimate of our main identification strategy. Lastly, we provide a discussion as to the implications of our results as they relate to the guiding research question.

A. Findings from Identification Strategy

As can be seen on Table 4, when running the instrumental variable regressions without fixed effects, the results observed are negative, which goes against our original intuition that having a pick in the draft lottery would actually increase winning percentage in the following year. When fixed effects are included in the model, we see the impact of a lottery pick on team winning percentage decreases even further, which again goes against our original beliefs on the matter. Contextualizing these results, this would mean that if a team has a pick in the draft lottery, then in the next season, we would expect their winning percentage to decrease by the varying levels shown in Table 4. Initially, this may make sense if we believe all rookies to have a negligible impact (or possibly a negative one) on a team in their first season in the league, which may certainly be true, even though we generally expect rookies drafted in the lottery to be excellent players capable of making an impact.

In any case, the effect of having a pick in the lottery on team winning percentage in the following year is not statistically different from zero for each of the different regressions shown in Table 4. A possible reason that we observe such levels of significance is due to the general lack of data that is available for further study about the NBA, as only a certain number of seasons have actually been played since the draft lottery was instituted. Since this problem is intrinsic to the research question in relation to the NBA, we realize this is a potential limitation of the model we have decided to follow in looking for an answer. However, we do assume in continuation that the statistical insignificance of the results is valid, since our methodology is correct with the data available. We will address this again in the discussion about checks for robustness.

Though statistically insignificant, there are potentially important economic repercussions that follow from this finding. In essence, since drafting in the lottery has no statistical impact on a teams' winning percentage for the following season, it seems to be the case that the tanking strategy does not seem to be effective; however, we make further implications in the discussion that soon follows.

B. Checks for Robustness

To ensure that the results we found are as accurate as they can be, we performed different robustness checks. These checks help to reinforce the claims we make with our main regression by showing similar results with an expanded bandwidth on our main regression, and by showing opposite results when considering the effect of making the playoffs rather than being in the lottery.

In the first step of checking for robustness, we ran our main regression with larger bandwidths to check if no effects hold across larger bandwidths. As shown in Table 5, there is no significant effect up to a bandwidth including conference ranks from four to twelve. However, larger bandwidths yield significant negative effects on future winning percentage. These results are most likely not of great concern for two reasons. One, there could be randomness where results are significant because at the one percent significance level, one would expect one out every 100 tests to be significant even when they are not, and considering that no other result was significant, that could be the case here. The stronger argument against the results being significant is that the larger the bandwidth, the more likely selection bias affects the results. With a small bandwidth, the only differences between the teams is that some made the playoffs and others missed them and that difference was as good as random, as we discussed earlier. With a larger bandwidth, there are more inherent differences between teams that were at the top of the conference and those at the bottom. This selection bias mentioned is most likely the reason for the negative effect of being in the lottery that we end up observing in the larger bandwidth.

Another robustness check we performed was to look at the effect of making the playoffs on future winning percentage, as seen in Table 3 and Figure 3. This checks the opposite effect as teams that were not in the lottery mostly made the playoffs. Here we found positive, but again insignificant results, as can be mostly seen again in Figure 3. These results conform with the notion that making the playoffs helps a team in the following year for various reasons. We again express the notion that these results, with more data generally available, could be statistically significant as the standard errors are small and approach levels of significance. This is especially the case, in comparison with our main identification strategy, since the standard errors are even smaller and we get even closer to significance.

C. Implications

Due to the insignificant results, we find that a team should not purposefully miss the playoffs with the intention of improving through the draft. This is an important finding as many NBA executives are debating the merits of missing the playoffs in order to secure a lottery pick to increase future winning. This study shows that missing the playoffs on purpose is not beneficial in future years. Because making the playoffs will always be beneficial in the current year through playoff bonuses, extra attendance and broadcasting revenue, and prestige, we suggest that if a team is in the playoff race, that team should always try to make the playoffs.

While our results show that being in the lottery has no effect, it may be due to how we determined the model. One reason being in the lottery may not have an effect is the time it takes for draft picks to have a significant impact on the team. One extreme example is that of Lebron James. When the Cleveland Cavaliers drafted Lebron James, the team was not able to make the playoffs until his third season. Even for one of the most successful players of all time, his impact on the team was not significant until a few years later in his career. Thus, the impact of players selected by teams that are on the threshold of making the playoffs would likely take longer before it is actually felt, as it is probable that a rookie even with transcendent talent like James is unable to influence his team so greatly. In short, if being in the lottery does have an effect on future winning percentage, the effect would be more likely to appear several years after rather than in the following year.

V. Conclusion

We find that there is no difference in missing the playoffs and making the playoffs for teams on the edge of the playoff race in determining teams' winning percentage in the year following, nor is there a difference in being in the lottery or not. However, making the playoffs is more likely to have an effect on future winning percentage, at least in the year immediately following. An effect for being in the lottery may be more noticeable if we look at winning percentages further in the future. Studying past one year is difficult, however, due to the constant movement of players and coaches from team to team in free agency and through trades, as well as the player development from players already on the team. This results in many confounding variables that make it difficult to know the true effect of only missing the playoffs in one year. In spite of this difference, we still argue that our results hold.

In short, we conclude that there is no significant difference between the future success of lottery and non lottery teams close to the threshold of making the playoffs. Because of this finding, if a team is looking to be successful in the immediate future, then it is in that best team's interest to make the playoffs during the current season instead of trying to draft a talented rookie in the lottery. However, we repeat that our research and analysis still does not conclude definitively whether tanking for future success in the long-term is effective and invite further research to be done on the matter, if plausible and practical.

Works Cited

Cohen, Josh. “The History of No. 1 vs. No. 8 in the NBA Playoffs.” *Orlando Magic*, NBA.com/Magic, 10 Aug. 2020, www.nba.com/magic/history-no-1-vs-no-8-nba-playoffs-20200810.

Gerchak, Yigal, et al. “The Evolution of Draft Lotteries in Professional Sports: Back to Moral Hazard?” *Interfaces*, vol. 25, no. 6, 1995, pp. 30–38. *JSTOR*, www.jstor.org/stable/25062073. Accessed 9 Dec. 2020.

Tim Reynolds | The Associated Press. “10 Things to Know as 2020 NBA Playoffs Begin.” NBA.com, NBA.com, 10 Oct. 2020, www.nba.com/news/10-things-know-playoffs-begin.

Soebbing, Brian & Humphreys, Brad. (2010). Do Gamblers Think That Teams Tank? Evidence from the NBA. University of Alberta, Department of Economics, Working Papers. 31. 10.1111/j.1465-7287.2011.00298.x.

“Land Of Basketball.com.” *NBA 1984-85 Regular Season Standings*, www.landofbasketball.com/yearbyyear/1984_1985_standings.htm.

“NBA Salaries - Dataset by Datadavis.” *Data.world*, 11 June 2019, data.world/datadavis/nba-salaries.

“NBA Team Continuity ('52-'16) - Dataset by Gmoney.” *Data.world*, 10 Jan. 2017, data.world/gmoney/nba-team-continuity-52-16.

“NBA Team Records by Year - Dataset by Gmoney.” *Data.world*, 24 Jan. 2017, data.world/gmoney/nba-team-records-by-year.

“1976 To 2015 Nba Draft Data - Dataset by rvino88.” *Data.world*, 11 Oct. 2016, data.world/rvino88/1976-to-2015-nba-draft-data.

Appendix

Table 1: Summary Statistics

Variable	Obs	Mean	Std. Dev.	Min	Max
Year	1,168	1999.277	9.622619	1982	2015
Total Salary	1,122	4.26E+07	2.94E+07	0	1.62E+08
Conference	1,122	0.504456	0.500203	0	1
Conference Rank	1,122	-0.15062	4.136125	-7	7
Games Behind 8 Seed	1,122	1.100267	13.04352	-37	35
Make Playoffs	1,122	0.547237	0.497986	0	1
Draft Position	938	15.60341	8.643075	1	30
Future Win Percentage	1,164	0.494886	0.153972	0.106	0.89
TeamContinuity	1,113	0.669371	0.162708	0.11	0.99

Table 2: Balance of Covariates

Covariates	(1) Salary	(2) Continuity
	-1.762e+06	-0.00184
	(3.155e+06)	(0.0152)
Observations	399	399
R-squared	0.007	0.010

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Figure 1:

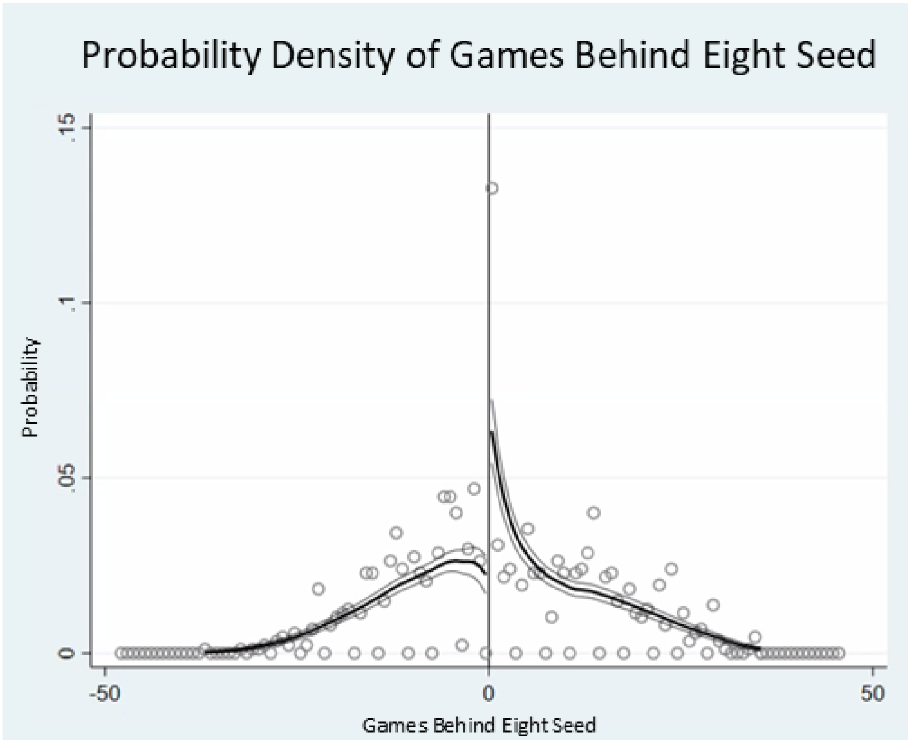

Figure 2a:

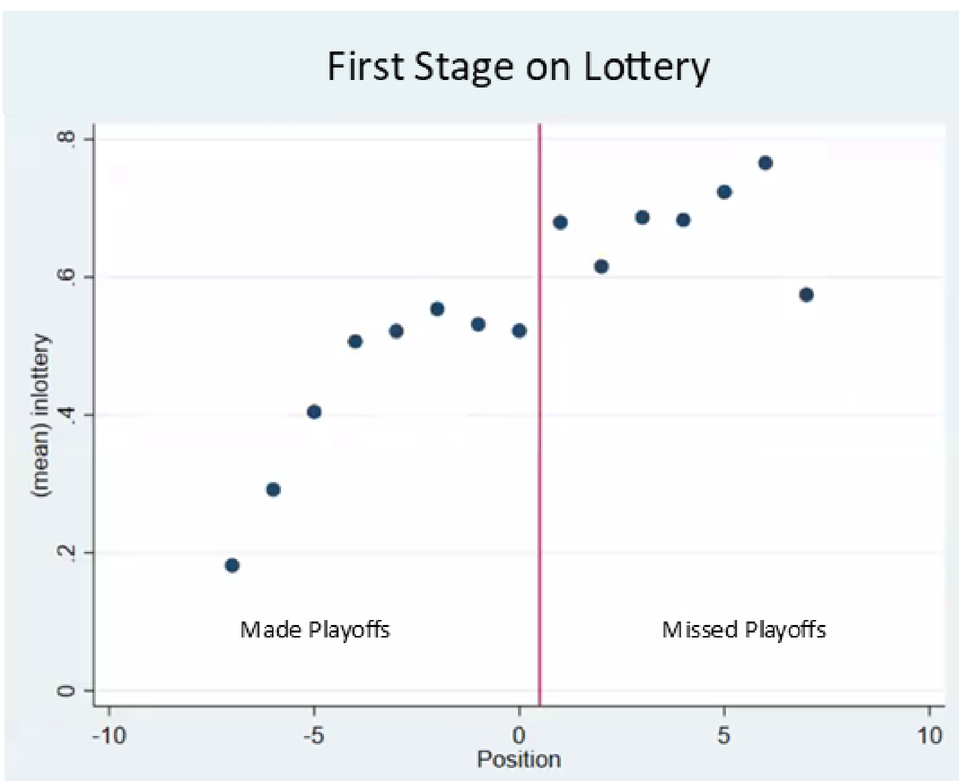

Figure 2b:

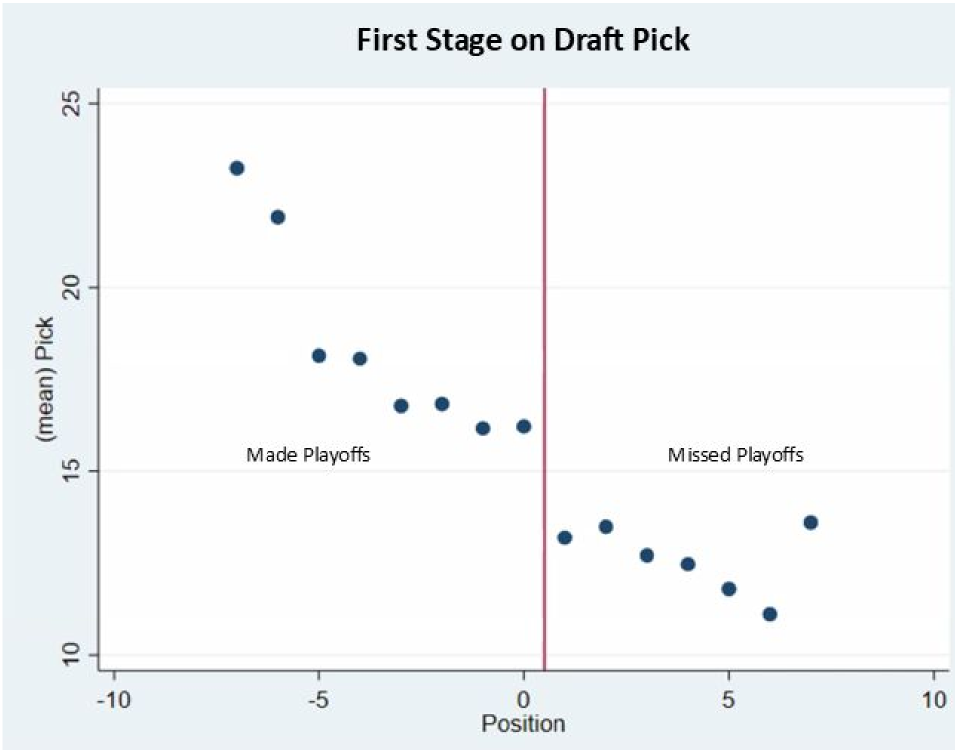

Figure 3:

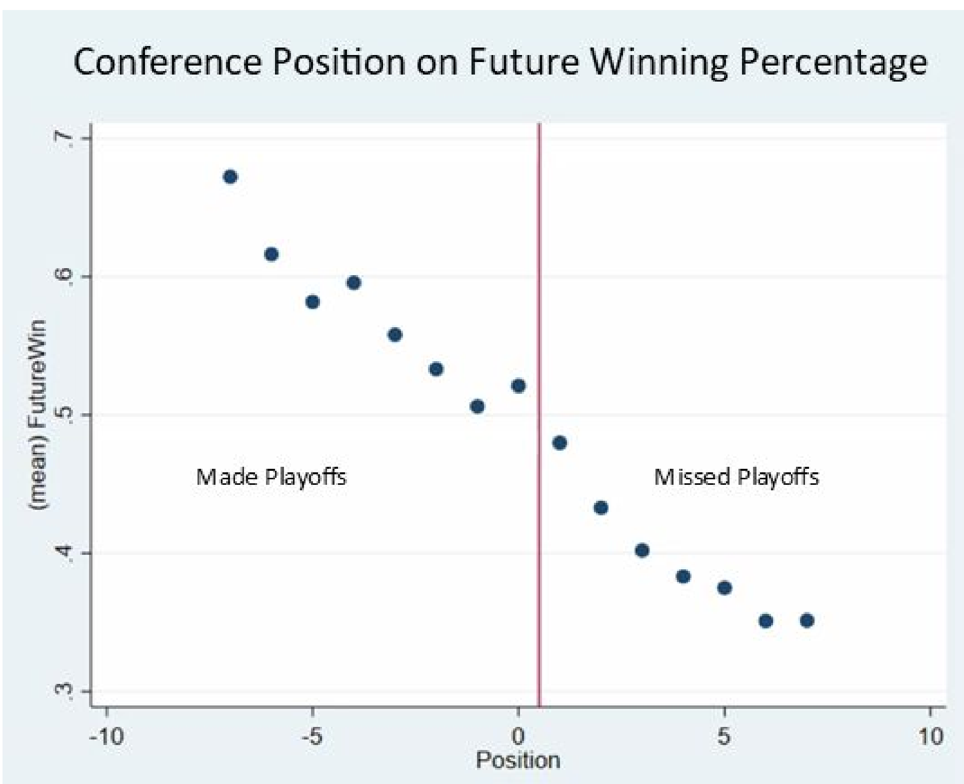

Table 3:

Effect of Making the Playoffs on Future Winning Percentage			
	(1)	(2)	(3)
	Model 1	Model 2	Model 3
	0.0410	0.0400	0.0403
	(0.0268)	(0.0269)	(0.0266)
Controls?	No	Yes	Yes
Fixed Effects?	No	No	Yes
Observations	398	398	398
R-squared	0.063	0.065	0.232

Notes: this table contains regression discontinuity-based estimates of the effect of being making the playoffs on future winning percentage. Each model includes only observations with teams that have Conference Positions between 6 and 10. Controls include total team salary and team continuity. Teams are included in fixed effects

Robust standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Table 4:

Effect of Being in the Lottery on Future Winning Percentage with Instrumental Variable			
	(1)	(2)	(3)
	Model 1	Model 2	Model 3
	-0.674	-0.685	-4.703
	(0.725)	(0.773)	(30.15)
Controls?	No	Yes	Yes
Fixed Effects?	No	No	Yes
Observations	398	398	398

Notes: this table contains instrumental regression discontinuity-based estimates of the effect of being in the lottery on future winning percentage. Each model includes only observations with teams that have Conference Positions between 6 and 10. Controls include total team salary and team continuity. Teams are included in fixed effects

Robust standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Table 5:

Effect of Being in the Lottery on Future Winning Percentage at Different Bandwidths				
	(1) Model 1	(2) Model 2	(3) Model 3	(4) Model 4
	-4.702 (30.151)	-1.256 (1.283)	-0.4643*** (0.1121)	-0.5171*** (0. 10150)
Controls?	Yes	Yes	Yes	Yes
Fixed Effects?	Yes	Yes	Yes	Yes
Bandwidth	Conference Position = [6, 10]	Conference Position = [4, 12]	Conference Position = [2, 14]	All data
Observations	398	704	998	1,110

Notes: this table contains instrumental regression discontinuity-based estimates of the effect of being in the lottery on future winning percentage. Each model includes a different bandwidth of conference position. Controls include total team salary and team continuity. Teams are included in fixed effects

Robust standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1