Referee Bias in March Madness

Michael Lyons

Department of Economics/Accounting Honors Thesis

Advisor: Professor Robert Baumann

April 2019

ABSTRACT

The Men's Division I National Collegiate Athletic Association (NCAA) March Madness tournament generates 85% of the NCAA's revenue in a given year. This study examines referee bias in this tournament using data from the 2016-2017 and 2017-2018 seasons. Results indicate that referees do not call more fouls upon Blueblood teams. However, more fouls are called on teams that are seeded lower than their opponents and on teams who are leading in games. This study suggests that bias exists towards keeping games close and having highly ranked teams continue through to the final rounds in the tournament. These forms of bias could positively affect consumer demand for the tournament, the NCAA's primary form of revenue each year.

1. INTRODUCTION

March Madness has become a national phenomenon. Sixty-eight college basketball teams from across the United States battle to determine the national champion of Men's Division I College Basketball. From an economic standpoint, this tournament generates more than 85 percent of the National Collegiate Athletic Association's (NCAA) revenue in a given year (Schoettle, 2016). The vast majority of this income is from television and other media rights to broadcast the games to viewers in the United States and internationally (Schoettle, 2016). The NCAA has recently renewed its contract with Turner Sports (CBS, TNT, TBS, TrueTV) through 2032; the average income for this renewal is $1.1 billion annually (Battaglio, 2016). More than half of this income is distributed to member institutions, including large public universities and smaller private colleges (Schoettle, 2016).

The increased popularity of the tournament over the past two decades is boosted by the rising number of cable television stations and internet streaming venues, as well as the extensive marketing of gambling pools, focusing on the March Madness tournament bracket. Over the last decade, the television viewing audience has grown at a steady pace with some variations based on accessibility of the television station on which the game is broadcast (Paulsen, 2018). With the advent of internet streaming, individual game viewers have been more difficult to quantify, but overall viewership continues to increase on an annual basis.

The make-up of the Final Four, in terms of the presence or absence of historical powerhouse teams, is thought to have an effect on television rating and viewership (Bonesteel, 2018). Over the past two decades, championship games with historical national powerhouse teams, henceforth known as Bluebloods, have produced the largest television viewing audiences. Moreover, Blueblood games that are expected to be, and result in, close matches garner the

highest ratings of all tournament games (Bonesteel, 2018). As Sean McManus, Chairman of CBS Sports, recently stated with respect to the 2018 tournament, "from a television standpoint you really root for the big teams…Last night's Kentucky outcome [loss] was not good for us and not good for TBS at all. Kentucky being the bluebloods that they are…that really hurt us." (Bonesteel, 2018). This explicit emphasis on the importance of historical powerhouse teams in the tournament, as well as a desire to keep games close, raises concerns about bias during the games themselves.

A handful of studies examines referee bias in the National Basketball League (NBA) and college basketball and also provides a preliminary examination of the potential profitable nature of such biases (Thu, Hattman, Hutchinson, Lueken, Davis, and Linboom, 2002; Anderson and Pierce, 2009; Price, Remer, and Stone, 2012; Caudill, Mixon, and Wallace, 2014). The present study extends the extant literature on referee bias in college basketball in a number of ways. Rather than focus on regular season games from over ten years ago (Anderson and Pierce, 2009), I provide an analysis of NCAA tournament games by using data from the 2016-2017 and 2017-2018 March Madness tournaments. Furthermore, I introduce a new way to conceptualize bias that is possibly related to consumer demand. Since the NCAA tournament is played on neutral courts, the tournament offers a unique opportunity to examine contests without definitive home team advantage. Instead of home teams benefitting in contests, I test whether historical powerhouse teams, or Bluebloods, receive preferential treatment from the referees, even when they are not playing on their home court. In addition, since the current marquee status of a team may affect referee decision making, I examine a team's seeding in the March Madness tournament relative to their opponent in each game.

This study also extends the examination of close game bias in a context where every game is of the utmost importance: a single elimination tournament capable of ending a season in a dramatic fashion. Thus, a close-game bias, specifically the effect of leading in a game upon referee decision making, is investigated. Consistent with past research (Price et al., 2012), I also examine whether the magnitude of a score differential (trailing or leading by more than ten points) affects foul calling.

In summary, the present study examines referee bias in the Men's Division I NCAA March Madness tournament using data from the 2016-2017 and 2017-2018 seasons. Results indicate that referees do not call more fouls upon Blueblood teams during the NCAA tournament. However, more fouls are called on teams that are seeded lower than their opponents in March Madness games. More fouls are also called on leading teams. However, the magnitude of the score differential does not affect foul calling. This study suggests that bias exists towards keeping games close and having highly ranked teams continue through to the final rounds in the March Madness tournament. These forms of bias could positively affect consumer demand for the tournament, the NCAA's primary form of revenue each year.

2. LITERATURE REVIEW

Refereeing sports involves split second decisions that are prone to subjectivity and may have significant consequences related to the outcome. Because of this, referee bias has been studied extensively for over a decade (Dohmen and Sauermann, 2016). In all sports, referees are expected to be unbiased observers who make judgements based on an expert level understanding of the rules. One may assume that simply incentivizing referees to be impartial would be sufficient to ensure unbiased officiating. Numerous studies, however, suggest that outside factors

do indeed influence referees, pushing their decisions away from the ideal of unbiasedness (Garciano, Palacios-Huerto, and Prendergast, 2005; Mills, 2014; Snyder and Lopez, 2015).

One of the first studies on referee bias is Garciano, Palacios-Huerta, and Prendergast (2005). This study investigates bias in football (soccer) in regards to allowance of stoppage time at the end of soccer matches in the Primera Division in Spain. The authors find that end-of-game stoppage time is, on average, about 113 seconds longer when the home team is behind by one goal compared to when the home time is ahead by one goal (Garciano et al., 2005). They argue that the referees are incentivized to please the home supporters. There are comparable studies for different soccer leagues, including the German Bundesliga (Sutter and Kocher, 2004) and the English Premier League (Rickman and Witt, 2008). Even when controlling for the amount of yellow and red cards, substitutions, and injury treatments, there are significant differences in the amount of second-half stoppage time allotted to the home team in close games.

After these initial stoppage time studies, many examined other indicators of home-crowd referee bias including number of goals scored (Boyko, Boyko, and Boyko, 2007), number of penalties called (Dawson, Dobson, Goddard, and Wilson, 2007), and overall game outcome (Boeri and Severgnini, 2011). All of these studies show a statistically significant difference in outcome measures based upon home team advantage. Furthermore, researchers also investigated the determinants of a referee's biased behaviors. The importance of the contest (Garciano et al., 2005), the attendance and composition and the crowd (Dohmen and Sauermann, 2016), and the distance of the crowd from the field (Dawson and Dobson, 2010) all play pivotal roles in the degree of referee bias in soccer outcomes.

Beyond soccer, referee bias has been studied in other sports, both in the United States and internationally. Elements of bias, including home team advantage, favoritism based upon

nationality, and bias towards players in specific positions, have been identified in baseball, rugby, ice hockey, figure skating, American football and basketball (Pollard and Pollard, 2005). Relatively few studies look at the determinants of referee bias in basketball, either at the professional and the collegiate level. One of the most well-known, and controversial studies, investigates possible racial discrimination in the National Basketball Association (NBA) (Price and Wolfers, 2010). The authors find that NBA players of the same race as the official have four percent fewer fouls called against them than NBA players of a different race from the official. Additionally, players with the same race as the referee score 2.5 percent more points per game. Although Price and Wolfers (2010) has been challenged by the NBA, its results suggest that referee bias could be large enough to impact the outcome of some NBA games.

A more recent study builds upon the notion that "many NBA fans have long been aware that NBA referees appear to be biased, particularly with respect to their treatment of star players" (Caudill et al., 2014). The authors focus on the NBA Playoffs, due to the importance of the games, to examine the number of extra fouls given to All-Star players at the end of the fourth quarter in comparison to the number of fouls given to other NBA players. They find that All-Star players have more free throws at the end of playoff contests than non-All-Star players. They also show evidence of a close-game bias, with more fouls called against the team with the lead. These data suggest that referee bias might revolve upon a desire, implicitly or explicitly, to create exciting contests for viewers – ones where the stars shine and the score remains close. The authors state that these incentives could be "an attempt to protect the league's main brand and increase consumer demand" (Caudill et al., 2014, 251).

A more direct examination of referee bias toward consumer demand and profitability in the NBA is by Price, Remer and Stone (2012). This research looks at a range of fouls and

turnovers over five years of NBA regular season and playoff games, and finds evidence that teams playing at home have more than an eleven percent advantage in turnovers and an eight percent advantage in shooting fouls. It also determines that a close-game bias exists, particularly for teams who are trailing by more than ten points. These trailing teams commit significantly fewer fouls, both shooting and non-shooting, and receive the advantage in turnovers called by the officials. Additionally, they find that trailing teams in a playoff series receive more favorable turnover calls than leading teams. They also show that the magnitude of the score differential affects referee behavior, with teams winning by more than ten points experiencing the greatest number of infractions against them. The authors conclude that these biases may positively increase consumer demand for the league.

Turning to college basketball, the literature on referee bias in NCAA men's basketball is rather scant. Two studies on referee bias in the NCAA discuss the notion of fair play among officials. Anderson and Pierce (2009) illustrate that officials attempt to equalize the number of fouls between teams in order to create the perception of fair, unbiased officiating. They also show evidence of a close-game bias, with more fouls being called on leading teams. Their data is the 2004-2005 regular season, but the study also includes games from the March Madness tournament that season. In an earlier study of fair play, researchers examine the 1999-2000 Men's Division I season for evidence of a close game bias (Thu, Hattman, Hutchinson, Lueken, Davis, and Linboom, 2002). These authors find that, when a game is nationally televised, the referees are more likely to call a higher number of fouls on the leading team.

3. DATA AND METHODOLOGY

3.1 Data

This study uses play-by-play data posted on espn.com from the 2016-2017 and 2017-2018 Men's Division I March Madness Tournaments. The espn.com data include the first four play-in games through the National Championship game, meaning that there is a total of 67 games in each tournament data set. Given the interest in assessing whether or not occurrences of fouls are affected by the Blueblood status of a team, the data set includes only games involving a Blueblood team. For this study, Blueblood teams are teams with eight or more Final Four appearances between 1939 and 2018. These teams, ranked by number of appearances, are University of North Carolina (20), University of California – Los Angeles (UCLA) (17), University of Kentucky (17), Duke University (16), University of Kansas (15), Ohio State University (10), Michigan State University (9), University of Louisville (8), and Indiana University (8). Data regarding Blueblood status are obtained through ncaa.com.

Consistent with previous literature in this area (Price, Remer, and Stone, 2012), the sample includes observations for each minute and for each team. For each minute per team, the number of personal fouls is recorded. Additionally, the data include whether a particular team is leading or trailing, and the magnitude of the lead or deficit at each minute. Since losing teams tend to commit intentional fouls during the final minutes of a game in order to stop the time clock, the data exclude observations from the last three minutes of the second half of each game.

In total, there are 2,604 observations from 35 games of the 2016-17 and 2017-18 NCAA March Madness Tournaments. There are 1,079 personal fouls over the course of these 35 games, and each team included in the data set commits an average of 15.25 personal fouls per game.

Data regarding other key variables are from ncaa.com. Specifically, the tournament seeding for each team and the round of play (Final Four or Championship) for each game are included from ncaa.com.

3.2 Methodology

Building on prior research in this area (Anderson and Price, 2009; Price, Remer, and Stone, 2012), this study utilizes linear regression models to test whether the number of fouls is affected by the team's Blueblood status and their seed in the tournament in comparison to its opponent. The study also assesses whether or not the magnitude of a lead impacts foul calling.

The initial regression model is as follows for game *i* and team *t*:

$$Fouls_{it} = \beta_0 + \beta_1 * Blueblood_i + \beta_2 * leading_{it} + \beta_3 * scorediffpos_{it} + \beta_4 * scorediffneg_{it} + \beta_5 * seeding_i + \beta_6 * attendance_i + \beta_7 * averageoffouls_i + \beta_8 * finalfours_i + e_{it}$$

In the model, the dependent variable is the number of fouls committed during each observation. The independent variables include *Blueblood*, a binary variable that equals one if the team is designated as a Blueblood team. Leading or trailing status and score are incorporated into three variables. *Leading* is a binary variable that equals one if the team is leading at the time of the observation and zero otherwise. *Scorediffpos* is also a binary variable that equals one if the team is winning by greater than ten points at the time of observation and zero otherwise. Likewise, *Scorediffneg* is a binary variable that is one if the team is losing by more than ten points at the time of observation and zero otherwise.

In order to examine the possibility that the current ranking and status (and thus not only historical Blueblood status) of the team may affect foul calls, the independent variables include *seeding*, which equals one if a team is ranked higher in the tournament than the team that they

are playing and zero otherwise. The model also includes an additional team level control variable, *averageoffouls*, which is the average number fouls committed by each team during the regular season. This variable takes into consideration the idea that some teams may foul more or less than other teams by nature.

Additionally, even though NCAA March Madness games are held on neutral courts, the intensity of the crowd may influence referee decision making. Specifically, as rounds progress, the intensity of the crowd is likely to increase. Thus, the round in the tournament, *Finalfours*, serves as an additional control variable. *Finalfours* is a binary variable that equals one if the tournament round is Final Four or the Championship game and zero otherwise.

Table 1 presents the means for the aforementioned independent variables. Standard deviations are included for those variables that are non-binary.

Table 1. Summary statistics

Variable	Mean (Std. Dev.)
Blueblood	0.57
Seeding	0.5
Leading	0.46
> 10 point lead	0.15
> 10 point deficit	0.15
Average Reg. Season Fouls	17.41 (1.54)
Final Four or Further	0.086

4. RESULTS

Table 2 reports the effects of several variables upon the number of fouls called during March Madness tournament games. The variables included in the linear model are as follows: a team's Blueblood status; their seeding in the tournament relative to their opponent; whether the

team is leading at the time of a foul; the magnitude of the lead or deficit; the average number of fouls committed by the team during the regular season; and the round in the tournament (Final Four or Championship game).

Table 2. Effects of Blueblood status, seeding, leading status, and magnitude of lead upon foul calls in the NCAA March Madness tournament

Variable	Coefficient	Std. Err.	t	*P*
Constant	-0.1009	0.1437	-0.7	0.483
Blueblood	-0.0394	0.0371	-1.06	0.289
Seeding	-0.0166	0.0377	-0.44	0.661
Leading	0.0929	0.0298	3.12	0.002
> 10 point lead	0.0243	0.0389	0.62	0.532
> 10 point deficit	0.0384	0.037	1.04	0.299
Average Reg. Season Fouls	0.0279	0.0081	3.46	0.001
Final Four or Championship	0.0674	0.0456	1.48	0.14
Tests of overall significance:				
F (7, 2585) = 4.27				
P = 0.0001				
R-squared = 0.0102				

The first hypothesis states that more fouls are called against non-Blueblood teams during the NCAA March Madness tournament because of the possible bias towards Blueblood teams, which are historical powerhouse teams with high marquee status. The results of the linear regression are not consistent with this hypothesis. The t-statistic is not significant ($p = 0.289$), and thus the coefficient for Blueblood does not differ significantly from zero. Therefore, Blueblood status does not affect the number of fouls called by referees during tournament games. Furthermore, the current seeding of a team in the NCAA March Madness tournament also does

not affect the number of fouls called by referees during tournament games. The *p*-value for seeding is not outside typical significance ranges ($p = 0.66$), and thus the coefficient for seeding does not differ significantly from zero. Thus, the results are not consistent with the hypotheses that higher seeds or Blueblood teams would receive more fouls during NCAA March Madness tournament games.

The third hypothesis states that more fouls are called against teams leading during games in the NCAA March Madness tournament. The results of the linear regression are consistent with this hypothesis. The coefficient for the leading variable is both positive (0.0929) and statistically significant ($p = 0.002$), meaning that teams with the lead during tournament games are more likely to have calls made against them. This result is significant even when controlling for the average number of fouls committed by a team during the regular season and the intensity of a game as measured by the round of the game in the tournament (Final Four or higher), suggesting that foul calling during the tournament differs from the regular season.

The fourth hypothesis states that the magnitude of the lead influences referee foul calling. The results of this regression are not consistent with this hypothesis. In total, 806 of the 2,590 observations involved a lead or a deficit that was greater than ten points. The *p*-values for leads or deficits larger than ten points are not significant ($p = 0.532$ and $p = 0.299$ respectively), and thus the coefficients for these variables are not significantly different from zero. Thus, the magnitude of the lead does not affect foul calling during NCAA March Madness tournament games.

Since many Blueblood teams are also highly seeded teams, I estimate a second regression without the Blueblood variable. This provides a single independent measure of the effects of a team's status upon referee foul calling: seeding. I also remove the other insignificant independent

variables examining the magnitude of the score differential from the second linear regression model. Table 2 reports the data examining the effects of leading status and seeding while controlling for the average number of fouls per team during the regular season and the round of each game in the tournament (Final Four or Championship game).

Table 3. Effects of seeding and leading status upon the number of fouls called in the NCAA March Madness tournament

Variable	Coefficient	Std. Err.	t	*P*
Constant	-0.0434	0.1535	-0.28	0.777
Seeding	-0.0542	0.027	-2.01	0.045
Leading	0.0835	0.0264	3.16	0.002
Average Reg. Season Fouls	0.0252	0.0085	2.95	0.003
Final Four or Championship	0.1126	0.0637	1.77	0.077
Tests of overall significance:				
F (4,2585) = 6.24				
P = 0.0001				
R-squared = 0.0084				

In the second linear regression, both independent variables are significant. Specifically, the coefficient for seeding is negative (-0.0542) and significant (p =0. 045). Thus, lower seeded teams receive more fouls than their higher seeded opponents do during NCAA March Madness tournament games. The effects of leading are similar to the previous estimation; the coefficient is positive (0.0835) and significant (p = 0.002). Therefore, teams with the lead are more likely to have more fouls called against them during tournament games.

It is important to note that the average number of fouls committed by teams during the regular season also affects referee decision making (p = 0.003). The number of fouls called on

teams in the data during the regular season ranges from 15.03 to 20.97 and the average is 17.41. Furthermore, the round of the tournament also approaches significance in the model ($p = 0.077$). The independent variables of seeding and leading status are significant even when controlling for the average number of fouls committed by teams during the regular season and the round of play in the March Madness tournament (Final Four or Championship game).

5. DISCUSSION

The results of the present study suggest that Bluebloods, or historical powerhouse teams, do not receive preferential treatment from the referees in the NCAA March Madness tournament. However, current marquee status, as measured by a team's seeding in the tournament relative to their opponent, affects referee decision making as higher seeded teams are less likely to receive foul calls than their lower seeded opponents. In a tournament without a true home court advantage, having higher ranked teams proceed through to the final rounds of the tournament could be beneficial for the NCAA. The momentum that these teams have developed throughout the season could positively impact television ratings and the income generated from the tournament. Future research is needed to further investigate this idea.

Consistent with previous research in this area (Thu et al., 2002; Anderson and Pierce, 2009; Price et al., 2012; Caudill et al., 2014), the present study provides evidence of a close-game bias. Similar to the results of earlier investigations of NBA and NCAA regular season play, the present findings indicate that leading teams are more likely to receive foul calls by referees in the March Madness tournament. However, unlike the past research by Price et al. (2012), this study suggests that the magnitude of the lead does not affect referee decision making. It is possible that the present measure of the magnitude of a lead or deficit (more than 10 points) does not fully capture this variable. Future studies could examine the magnitude of a lead or deficit in

a more refined way. Overall, keeping NCAA March Madness games close in score could have important effects upon consumer demand for the tournament and television ratings.

There are a few key limitations to the present study. The sample is restricted to games from the 2016-2017 and 2017-2018 March Madness tournaments that include Blueblood teams. In order to fully investigate the seeding and close-game biases that have emerged during this study, further analyses are needed with the complete set of tournament games. This larger sample would also be important in order to assess the effects of game-level variables, such as round in the tournament (which approached significance in the present study), upon referee decision making. Additionally, the R-squared in the estimations are quite small, R-squared = 0.0102 and 0.0084, respectively. Future research could attempt to include a larger set of independent variables in order to increase the precision of the model.

Overall, the present study extends the literature on referee bias in NCAA basketball by providing an analysis of recent March Madness tournaments. The findings indicate that bias exists in referee decision making towards keeping games close and helping teams with strong regular seasons succeed.

REFERENCES

Anderson, Kyle J., and Pierce, David A. 2009. Officiating Bias: The Effect of Foul Differential on Foul Calls in NCAA Basketball. *Journal of Sports Sciences* 27 (7): 687.

Battaglio, Stephen. 2016. March Madness through 2032: CBS and Turner Extend Contract for NCAA Men's Basketball Tournament. *Los Angeles Times*, April 12.

Boeri, Tito, and Severgnini, Battista. 2011. Match Rigging and the Career Concerns of Referees. *Labour Economics* 18 (3): 349-359.

Bonesteel, Matt. 2018. Everyone Loves a March Madness Cinderella...Except for Television Networks. *Washington Post*, March 26.

Boyko, Ryan H., Boyko, Adam R., and Boyko, Mark G. 2007. Referee Bias Contributes to Home Advantage in English Premiership Football. *Journal of Sports Sciences* 25 (11): 1185-1194.

Caudill, Steven B., Mixon, Franklin G., Jr., and Wallace, Scott. 2014. Life on the Red Carpet: Star Players and Referee Bias in the National Basketball Association. *International Journal of the Economics of Business* 21 (2): 245-253.

Dawson, Peter, and Dobson, Stephen. 2010. The Influence of Social Pressure and Nationality on Individual Decisions: Evidence from the Behaviour of Referees. *Journal of Economic Psychology* 31 (2): 181-191.

Dawson, Peter, Dobson, Stephen, Goddard, John, and Wilson. John. 2007. Are Football Referees really Biased and Inconsistent?: Evidence on the Incidence of Disciplinary Sanction in the English Premier League. *Journal of the Royal Statistical Society. Series A (Statistics in Society)* 170 (1): 231-250.

Dohmen, T. J., and Sauermann, J. 2016. Referee Bias. *Journal of Economic Surveys* 30 (4): 679-695.

Garciano, Luis, Palacios-Huerta, Ignacio, and Prendergast, Canice. 2005. Favoritism Under Social Pressure. *The Review of Economics and Statistics* 87 (2): 208-216.

Mills, Brian M. 2014. Social Pressure at the Plate: Inequality Aversion, Status, and Mere Exposure. *Managerial and Decision Economics* 35 (6): 387-403.

Paulsen. 2018. NCAA Tournament Down from Last Year, Up from '16. Sports Media Watch., last modified April 4.

Pollard, R., and Pollard, G. 2005. Long-Term Trends in Home Advantage in Professional Team Sports in North America and England (1876-2003). *Journal of Sports Sciences* 23 (4): 337-350.

Price, Joseph, Remer, Marc, and Stone, Daniel F. 2012. Subperfect Game: Profitable Biases of NBA Referees. *Journal of Economics & Management Strategy* 21 (1): 271-300.

Price, Joseph, and Wolfers, Justin. 2010. Racial Discrimination among NBA Referees. *The Quarterly Journal of Economics* 125 (4): 1859-1887.

Rickman, Neil, and Witt, Robert. 2008. Favouritism and Financial Incentives. *Economica* 75: 296-309.

Schoettle, Anthony. 2016. NCAA Wants New TV Deal: March Madness Contract is 85 Percent of Revenue. *Indianapolis Business Journal,* January 11.

Snyder, Kevin, and Lopez, Michael. 2015. Consistency, Accuracy, and Fairness: A Study of Discretionary Penalties in the NFL. *Journal of Quantitative Analysis in Sports* 11 (4): 219-230.

Sutter, Matthias, and Kocher, Martin G. 2004. Favoritism of Agents – the Case of Referees' Home Bias. *Journal of Economic Psychology* 25 (4): 461-469.

Thu, Kendall M., Hattman, Kelly, Hutchinson, Vance, Lueken, Scott, Davis, Nathan, and Linboom, Elmer. 2002. Keeping the Game Close: "Fair Play" among Men's College Basketball Referees. *Human Organization* 61 (1): 1-8.