

THE UNIVERSITY OF CHICAGO

ARE NBA COACHES TOO CONSERVATIVE WITH PLAYERS IN FOUL TROUBLE?

A DISSERTATION SUBMITTED TO

THE FACULTY OF THE UNIVERSITY OF CHICAGO

BOOTH SCHOOL OF BUSINESS

IN CANDIDACY FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

BY

DANIEL K. WALCO

CHICAGO, ILLINOIS

JUNE 2018

Copyright © 2018

Daniel K. Walco

Table of Contents

List of Tables	iv
List of Figures	v
Acknowledgements.....	vi
Abstract.....	vii
Introduction.....	1
Description of Data.....	9
Section 1: How does foul trouble affect playing time?.....	10
Data	11
Analysis.....	12
Discussion.....	19
Section 2: Does foul trouble affect performance?	19
Data	20
Analysis.....	25
Discussion.....	35
Section 3: Is the end of the game meaningfully different from the rest of the game?.....	35
Data	36
Analysis.....	37
Discussion.....	38
General Discussion	39
References.....	49
Appendix 1: List of star players.....	52
Appendix 2: Offensive rating details	53
Appendix 3: Defensive rating details.....	55
Appendix 4: Section 2 robustness checks	56
Appendix 5: Effect of players' skill difference on outcome, by game segment	58
Appendix 6: Foul out simulation details	59

List of Tables

Table 1: Regression of minutes played on minutes in foul trouble and VORP	15
Table 2: Regression of minutes played on minutes in foul trouble, star, and VORP	16
Table 3: Regression of minutes played on minutes in foul trouble, bucked by VORP	17
Table 4: Number of stints for each game segment, by foul trouble	21
Table 5: Regression of scoring margin on each team's foul trouble, with controls	27
Table 6: Regressions of team foul rates and team field goal percentages	31
Table 7: Regression of scoring margin controlling for foul rates and field goal percentages	32
Table 8: Regressions of foul rates, controlling for total fouls	33
Table 9: Cost of benching players in foul trouble.....	46
Table 10: Robustness checks for regression model in Section 2	56
Table 11: Additional Robustness checks for regression model in Section 2 (by game segment)	57
Table 12: Regression results of score differential on various measures of player quality, by game segment	58

List of Figures

Figure 1: Probability of playing in each minute, by number of fouls	12
Figure 2: Probability of playing in each minute, by number of fouls (low minutes per game).....	14
Figure 3: Distributions of score differential in each stint, by foul trouble	22
Figure 3a: Home player in foul trouble.....	22
Figure 3b: Away player in foul trouble.....	22
Figure 3c: No home player in foul trouble.....	22
Figure 3d: No away player in foul trouble.....	22
Figure 4: Stints with a player in foul trouble, by team	23
Figure 5: Coefficient for VORP.dif on scoring margin, by game segment	38

Acknowledgements

I would like to express my sincere gratitude to my co-chairs, Professors Jane Risen and George Wu, for their tireless help and support throughout this process. It has truly been a privilege to work with and learn from each of them over the last six years.

I would also like to thank my committee members, Professors Richard Thaler and Dan Bartels. The idea for this investigation came as a result of a conversation with Professor Thaler, whose insights and perspectives have been invaluable. And I will be forever grateful for the mentorship and guidance that Professor Bartels has provided throughout my graduate school career.

Abstract

Coaches in the National Basketball Association (NBA) typically bench players who are perceived to be in danger of fouling out. I examine the efficacy of this strategy. At a baseline level, it seems dubious to guarantee that a player misses playing time for fear that he might miss time later in the game. However, there are broadly two categories of reasons that coaching conventional wisdom might be optimal. First, it is possible that players who are in foul trouble tend to play poorly, and thereby hinder their team's performance. And second, the end of the game might be meaningfully different from the rest of the game, such that having the team's best players available for the final minutes is more valuable. Section 1 demonstrates that benching players in foul trouble does not merely shift the minutes that players would typically rest, but instead decreases their overall playing time. Section 2 reveals that having a player on the court in foul trouble actually improves team performance. And Section 3 provides evidence showing that the play at the end of games does not justify the decrease in playing time that accompanies benching foul-troubled players. Taken together, this analysis demonstrates that in general, coaches should not bench their players because of foul trouble.

Introduction

“Durant’s Foul Trouble Proves Costly to Thunder” (Thompson, 2012). That was the headline that the *New York Times* ran after Game 3 of the 2012 National Basketball Association (NBA) Finals matchup between the Oklahoma City Thunder and the Miami Heat. With 5:41 left in the third quarter, All-Star Kevin Durant was whistled for his fourth foul, at which point he was benched for the rest of the quarter. When he left the game, the Thunder led 60–54. By the time he re-entered, they trailed 71–67—a 17–7 run for the Heat in Durant’s absence. The Thunder ended up losing the game, and ultimately the series.

But was Durant’s foul trouble in itself really the culprit, as the *New York Times* headline suggested? Or was it instead the way in which Thunder coach Scott Brooks handled Durant’s foul trouble? NBA rules state that a player is not disqualified from playing until he commits six personal fouls. Yet Scott Brooks is not alone in pre-emptively benching players who appear to be in danger of reaching that threshold. Coaches typically bench players who have two fouls in the first quarter, three fouls in the second quarter, four fouls in the third quarter, or five fouls at any point; henceforth the “Q+1” strategy (Maymin, Maymin, & Shen, 2012).

In the present thesis, I investigate whether or not the Q+1 strategy is optimal. On its face, it seems that it is not. Let us assume for the sake of argument that coaches bench a player in foul trouble in order to reduce the chances of that player fouling out. That is, coaches are averse to the possibility that a player might be forced to miss playing time later in the game. I argue that it is counter productive to guarantee that a player misses some amount of time now for fear that he may possibly miss time later. That is akin to starving to death to ensure that you do not run out of food.

Why benching players in foul trouble might be optimal

Of course, this is oversimplifying the problem. The goal in basketball is not to maximize playing time for your best players, but rather to win the game. With this in mind, there are broadly two categories of reasons that benching players in foul trouble might be the optimal strategy. First, it is possible that players tend to play worse when they are in foul trouble than when they are not in foul trouble. Players are not robots; a player who knows that he is in danger of fouling out is likely to play less aggressively, which may be detrimental to his team. Relatedly, opponents might intentionally attack players in foul trouble, forcing them to either risk getting another foul or allow an easy basket. If there is in fact a drop in performance, and if that drop is sufficiently large, it may make sense to substitute the foul-troubled player out of the game.

Second, it is possible that not all minutes are the same, and that having a starter available at the end of a game is more important. This jibes with fans' intuitions regarding the decision, as I will discuss in more depth later. Stern's (1994) model of win probability shows that a point towards the end of a close game has a larger effect on a team's probability of winning than does a point towards the beginning of that game. Furthermore, Maymin, Maymin, and Shen (2012) use this concept of win probability in their analysis of foul trouble substitution decisions. They argue that the Q+1 rule is essentially optimal. However, because their model is based on win probability, which, as I described, naturally weighs events at the end of the game more than those at the beginning of the game, I believe they overestimate the effectiveness of benching players in foul trouble. Benching players early does generally ensure that they will be available later, and so if a model inappropriately overweights later minutes (as I contend theirs does), then benching players in foul trouble will appear to be more beneficial than it really is.

While I do not doubt the statistical truth that a point scored at the end of the game has a larger effect on win probability than one scored at the beginning of the game, I do question the logic. Qualitatively, the reason that a point scored in the first quarter has a small effect on win probability is because it is unknown whether or not that point will end up being meaningful. That is, if the game ends up being close, that point will have been very important, whereas if the game ends up being a blowout, that point will not have mattered at all. But obviously at the time it is scored, the outcome of the rest of the game is unknowable. Conversely, at the end of a game, it is known whether a point is important or not. Therefore, a point scored at the end of a close game is certainly meaningful, and therefore has a large effect on win probability, while a point scored at the end of a blowout has no impact at all. Looking at the totality of the game in retrospect, though, the timing of the points is irrelevant. A layup in the first quarter counts for two points, just as many as a layup in the fourth quarter. And an extra two points in the first quarter necessarily means that the team will have two more points in the fourth quarter than they otherwise would have had. The following stylized example illustrates this point.

Imagine that Team A's best player is Player X. For the purposes of this example, assume that a game consists of exactly 200 possessions (50 possessions per quarter), and that Player X will end up playing exactly 150 of those possessions (i.e three quarters). When Player X is in the game, Team A outscores its opponent by precisely 20 points per 100 possessions, and when he is not in the game, Team A gets outscored by precisely 40 points per 100 possessions. The coach of Team A is deciding whether to have Player X sit for the third quarter or sit for the fourth quarter. If Player X plays the first two quarters, sits for the third quarter, and plays the fourth, the score differential at the end of each quarter will be +10 after the first quarter, +20 after the second quarter, tied after the third quarter, and +10 to end the game. If instead, he plays the first three

quarters and sits for the fourth, the score differential at the end of each quarter will be +10 after the first quarter, +20 after the second quarter, +30 after the third quarter, and +10 to end the game.

In the first version, Player X is in the game for some very high leverage minutes. His team is tied to start the fourth quarter, and he helps lead them to a hard-fought 10-point victory. The points he contributes in these final minutes will have a substantial effect on his team's win probability. In the second version, there are no high leverage minutes. Team A had a 30-point lead to start the fourth quarter, and that lead never drops below 10. Player X does not even play in the fourth quarter, and so based on win probability, it will appear that he contributed significantly less than in the first version. But in both cases, the end result is the same: the team wins by 10 points.

Clearly real basketball is not as simple as the stylized example described above. Nonetheless, I believe that case serves as a useful baseline, and demonstrates why a model that uses win probability is not ideal. That said, it is still possible that the end of the game is different from the rest of the game in more substantive ways. For example, it is possible that the game is faster at the end, and so minutes played at the end of the game have more possessions, or that certain kinds of players are more valuable at the end of the game.

Relevant psychological factors

There are several other examples of suboptimal decision-making in the world of sports. And in fact, many of the contexts in which decision makers consistently err involve overly conservative decisions. Consider, for example, a situation in which an NBA team is trailing by two points with possession and only time enough for one final shot. While teams would statistically be better off attempting a three-point shot—either winning or losing immediately—

they choose to attempt a two-point shot much more often (Walker, Risen, Gilovich, & Thaler, 2018). In baseball, managers often choose to sacrifice bunt, moving a runner from first to second base, in part to avoid a potential double play. Extensive analyses have demonstrated that this is a flawed strategy (e.g. Tango, Lichtman, & Dolphin, 2006). And in the National Football League (NFL), coaches facing a short fourth down situation must decide whether to punt the ball, thereby giving up possession but improving their field position, or attempt to go for a first down. Coaches ultimately decide to punt the ball far more often than they should (Romer, 2006).

Given the high stakes associated with these decisions, the incentives for coaches to win, and the amount of information coaches have at their disposal, these kinds of errors might come as a surprise to traditional economists. Decades of research in behavioral science, though, help shed light on two distinct but related questions: (1) what factors contribute to producing these errant decisions in the first place? And (2) given that these situations arise frequently, what factors prevent decision-makers from correcting their mistakes?

Psychological factors that produce the errant decision

Walker et al. (2018) describe these kinds of decisions as evidence of “sudden-setback aversion”. When faced with a decision between a “fast” strategy that is more likely to result in ultimate success but also carries immediate downside risk, and a “slow” strategy that is suboptimal in the long-term but is less likely to fail immediately, people overwhelmingly choose the slow option. This pretty clearly maps onto the context of foul trouble. Keeping a foul-ridden player in the game carries the immediate risk of that player accumulating additional fouls. Benching the player, a slow strategy in that it removes the immediate risk, feels much safer. I suggest that, despite feeling safer, it may be worse in the long run.

The authors establish two mechanisms for this effect. First, they demonstrate the role of myopic loss aversion (Benartzi & Thaler, 1995). Combining ideas from loss aversion (Kahneman & Tversky, 1979; Tversky & Kahneman, 1992) and mental accounting (Kahneman & Tversky, 1984; Thaler, 1985), myopic loss aversion describes situations in which a person demonstrates a gain-loss asymmetry for a timeframe that is narrowly defined. In the context of basketball, the “correct” frame is the entirety of a game. Nonetheless, in these kinds of situations, coaches appear to be loss averse to the immediate outcome—having their foul-ridden player get called for additional fouls. The second mechanism for sudden-setback aversion is that people believe that choosing a fast strategy tempts fate, particularly when that strategy seems like an unnecessary risk (Risen & Gilovich, 2008). A coach may feel that he is tempting fate by keeping a player in foul trouble in the game, which would lead him to believe that the player is more likely than usual to be called for additional fouls.

The availability heuristic might also lead coaches to overestimate the probability of a player fouling out (Tverksy & Kahneman, 1973). The availability heuristic describes situations in which people make probability or frequency judgments by relying on the ease with which examples come to mind. For example, in one study, participants were asked to estimate the ratio of English words that start with the letter K to words that contain the letter K as the third letter. Because it is easier to think of words that start with the letter K than to think of words that have K as the third letter, the majority of participants estimated that there are more words that start with K. Bringing this discussion back to basketball, it is presumably very easy for coaches to imagine or remember a scenario in which a player fouled out, and this likely inflates their perceived probability of that outcome occurring.

Similarly, the availability heuristic could also cause coaches to overestimate the relative effect of a player fouling out (i.e. missing playing time at the end of the game) as compared to a player being benched when in foul trouble (i.e. missing playing time earlier in the game). It is easier to imagine losing a heartbreaking game in the final seconds because a star player has fouled out than it is to imagine losing a game in which a star player's absence in the second quarter led to an insurmountable deficit. Furthermore, organizational research shows that when making risky decisions, managers tend to overweight worst-case-scenarios (March & Shapira, 1987). Regardless of the actual probability of a player fouling out, the possibility of losing a player for the rest of the game looms large and can have an undue influence on decisions.

Psychological factors that prevent correction

Basketball games, and sports more generally, are unique decision-making domains because nearly identical situations arise fairly often. Not only do coaches have the opportunity to learn from their own experiences, but the structured nature of the sport allows them to learn from the experiences of others as well. It is therefore particularly surprising that coaches appear to have difficulty learning from their mistakes, and the mistakes of their peers, in this domain.

One factor that makes learning difficult in this context is the ambiguity of the feedback that coaches receive (Jennings, Amabile, & Ross, 1982). This one type of decision—whether or not to bench a player who is in foul trouble—is not made in isolation. Coaches make numerous other decisions over the course of a game, and the effect of each decision is further obscured by the noise in the environment. Benching a player in foul trouble does not guarantee a loss, and keeping him in the game does not guarantee a win. Without a structured statistical analysis, it would be virtually impossible to parse out the effects of any single decision.

A second reason that coaches might not correct this mistake is that they may rarely even entertain the possibility of keeping a foul-ridden player in the game. In order to learn that benching players in foul trouble is sub-optimal, a coach must first contemplate keeping that player in the game. Research shows that people typically do not consider a making a different decision unless the counterfactual is obvious, and that following conventional wisdom is unlikely to elicit counterfactual thinking (Kahneman & Miller, 1986). A coach who keeps a player in foul trouble in the game only to see him foul out minutes later is very likely to consider the counterfactual scenario in which he had benched that player. However, a coach who benches a player in foul trouble and has him available for the end of the game is much less likely to consider the alternative scenario. Relatedly, the anticipation of regret could play a powerful role here. Even if coaches think they might have a better chance of winning by keeping a player in the game, they might anticipate feeling especially regretful if they go against conventional wisdom and fail (Miller & Taylor, 1995).

Of course, this logic extends to the perceptions of fans as well. An economist would likely point out that coaches are not merely incentivized to win. If a decision that bucks conventional wisdom fails, it is more likely to draw the ire of spectators, and a coach's job security is certainly affected by public opinion. However, even if this were the sole cause of coaches' poor decision making (which I contend it is not), this still begs the question of why the general populace holds these beliefs. Fans certainly have less information than do coaches, but research shows that lacking information does not necessarily account for their faulty beliefs (Walco & Risen, 2017). In one study, football fans were asked whether they thought a coach should go for it or punt in a particular fourth-and-short situation. Participants were provided with calculated win probabilities indicating that the objectively superior strategy was to go for the first

down. Nonetheless, a sizable proportion of people who identified going for it as the rational strategy still said they would punt. Given that the information gap cannot completely account for fans' judgments, it seems reasonable to assume that the same psychological factors that I proposed affect coaches' decisions—myopic loss aversion, tempting fate, the availability heuristic, ambiguous feedback, and a failure to consider counterfactuals—fluence fans as well.

Overview of analysis

I am not the first person to suggest that NBA coaches take an irrational approach to foul trouble. Weinstein (2010a and 2010b) succinctly lays out a similar theoretical argument, as do Moskowitz and Wertheim (2011), who also provide some initial empirical evidence. The present thesis builds on these arguments, providing new approaches and a more thorough and comprehensive analysis. I use a multi-faceted strategy to address this problem. In Section 1, I provide descriptive statistics regarding the effects of foul trouble on playing time. Then in Section 2, I investigate the extent to which performance is affected by foul trouble, both on a team level and an individual level. And finally in Section 3, I analyze if and how the end of the game is different from the rest of the game. Ultimately, these various arms of my analysis all point to the same conclusion: coaches behave much too conservatively in the way they handle players in foul trouble.

Description of Data

The raw data for this analysis comes almost entirely from bigdataball.com. The data set includes play-by-play information for every NBA game played by every team in the 2015-2016

regular season¹. The following plays are included in the data set: shots, rebounds, fouls, free throws, turnovers, substitutions, timeouts, violations, jump balls, and starts and ends of periods. For any given play, the data includes the game in which the play occurred, the five players on the court for each team, the period of the game, the time left in the period, the score of the game, the player responsible for the play, and a verbal description of the event. For all turnovers, the data set indicates which player, if any, caused the turnover (e.g. a steal) and which player, if any, surrendered the turnover. For all shots, X and Y coordinates are given to indicate the precise location on the floor where the shot was taken. For made shots, there is information regarding who assisted the shot (or if there was no assist). And for missed shots, the data indicates whether the shot was blocked, and if so, by whom. With 30 teams playing an 82-game schedule, that yields 1,230 total games played by 476 unique players. I exclude all overtime minutes from the analysis, as those minutes are unique from the rest of the game and foul trouble is less meaningful in overtime. The resulting data set includes 233,958 possessions and 250,745 points scored. In each section, I will explain how I reshaped the play-by-play data in order to address each relevant question.

Section 1: How does foul trouble affect playing time?

The first set of relevant questions relate to the extent to which foul trouble affects playing time. This is equivalent to asking whether coaches really do bench players in foul trouble, a check of a key premise on which I base my argument that coaches bench players in foul trouble

¹ Throughout the paper, I limit my analysis to a single season, because in parts of the analysis, I use end-of-season statistics for individual players (e.g. value over replacement player). I do not believe this season is unique, and would expect all results to generalize to other seasons. I will discuss this further in the general discussion.

more than they should. Some might contend that when coaches bench their players in foul trouble, they are merely shifting the minutes that the players play, and are not decreasing the players' overall playing time. All players need some rest during a game, so perhaps coaches are giving the players their rest when they are in foul trouble, and compensating later in the game.

Data

For this section, the play-by-play data was reshaped such that it includes minute-by-minute foul data for every starter in every NBA game in the 2015-2016 season². For each player in each minute in each game, this data set indicates whether or not the player played for any part of that minute, and how many fouls the player committed (if any). In total, that yields 590,400 observations, with each observation representing one minute in one game for one player (1,230 games * 10 starters per game * 48 minutes per game). Using the number of fouls a player is awarded in each minute, I calculate his cumulative fouls for any given moment of a game. I also indicate whether or not a player is in foul trouble using the Q+1 rule of thumb that coaches typically use. Note that foul trouble is a binary distinction—a player is either in foul trouble or not in foul trouble.

When examining the precise effect of foul trouble on playing time, I use the raw play-by-play data to calculate playing time and time in foul trouble for each starter in each game. I also include each player's end-of-season value over replacement player (VORP), which is a metric designed to summarize overall player quality. VORP uses box score statistics to estimate the number of points a player is worth above or below a replacement level player per 100 possessions for each team, and it is weighted by the number of minutes a player plays. In the present sample, players' VORP has a mean of 0.63 and standard deviation of 1.35. The

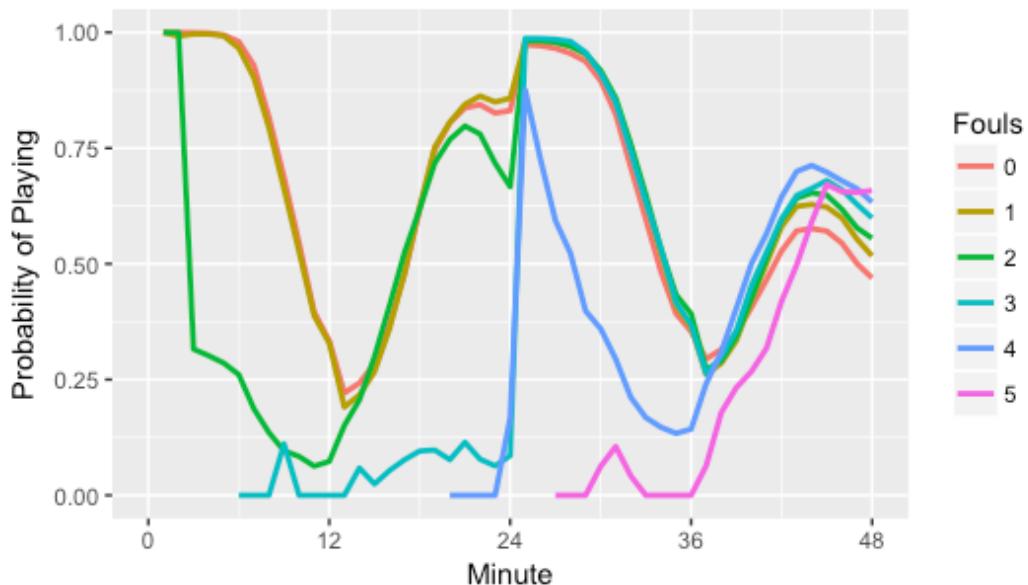
² This minute-by-minute data set was compiled by the Chicago Bulls.

distribution is heavily skewed, as only 13 players have a VORP over 4.0. Stephen Curry has the highest VORP of 9.8.

Analysis

First, for any given player-minute of a game, is a player less likely to be in the game if he is in foul trouble than if he is not? Figure 1 depicts the probability of a starter being on the court, separated by number of fouls, for each minute of the game.

Figure 1: Probability of playing in each minute, by number of fouls



The Q+1 rule is unmistakable. In the first quarter, starters with fewer than two fouls are treated identically, while starters with two or three fouls are much less likely to be in the game. For most of the second quarter, however, only starters with three or more fouls are treated differently. The minutes surrounding halftime present the starker evidence of Q+1 behavior. In minute 24, the last minute of the first half, starters with three fouls are only in the game 8.56% of the time. Just one game-minute later, in the first minute of the second half, starters are in the game 98.72% of the time. The Q+1 rule is also apparent in the third and fourth quarters, until the final few minutes of the game when foul trouble is no longer relevant. At that point, it appears that players

with more fouls are actually more likely to be in the game. But does that make up for the time missed earlier in the game?

In order to address that question, we can examine whether the number of minutes a player is in foul trouble negatively predicts his total playing time for the game. I find that it does. For each additional minute a player is in foul trouble, he plays approximately 0.27 fewer minutes (~16 fewer seconds) in the game, $b = -.272, t(11,925) = -21.515, p < .001$ ³. One may wonder if this is driven by players who foul out of the game. That is, perhaps players in foul trouble who end up fouling out account for the observed difference in playing time. If we exclude from the analysis all players who fouled out of the game ($n = 144$), we find that the result holds with a similar effect size, $b = -.282, t(11,781) = -21.321, p < .001$.⁴ Furthermore, this is not just due to substitutions made in games that end up being blowout. Looking at only those games that end with a final score differential less than 6 points, if anything, the relationship between foul trouble minutes and minutes played is even stronger, $b = -.303, t(3,341) = -12.711, p < .001$.

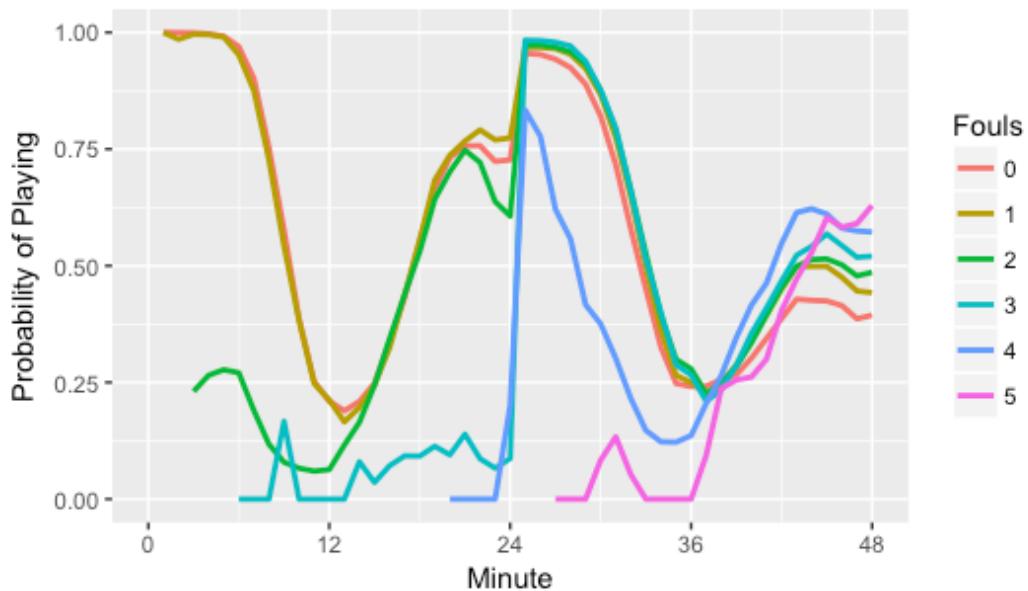
I similarly investigate whether being in foul trouble at any point in a game negatively predicts playing time. Based on the analyses described above, one might accurately assume that the answer once again, is yes. A player in foul trouble at any point in a game is predicted to play 1.54 fewer minutes in a given game, $b = -1.538, t(11,925) = -11.773, p < .001$. Again, limiting the analysis to players who did not foul out, I find the same result, $b = -1.638, t(11,781) = -12.219, p < .001$.

³ With 10 starters per game and 1,230 total games, there are 12,300 observations. Degrees of freedom are reduced by including fixed effects for players.

⁴ Of course, more players would foul out if they were not benched. I will address this in the General Discussion and in Appendix 6.

Thus far, my analyses indicate that on the whole, foul trouble has a negative effect on playing time. But is this effect the same for all players? For instance, players who tend not to play as many minutes in any given game are much less likely to foul out. They simply have less time to accumulate additional fouls. Nonetheless, coaches still bench these players when they are in foul trouble (see Figure 2).

Figure 2: Probability of playing in each minute, by number of fouls (fewer than 30 minutes per game)⁵



If these substitutions are made with an eye towards saving players for the end of the game, this pattern is difficult to reconcile. To better illustrate this point, consider two players who each have three fouls in the first half. One of these players (Player A) averages 26 minutes per game, and is therefore expected to play approximately 13 minutes in the second half. The other (Player B) averages 34 minutes per game, and is therefore expected to play approximately 17 minutes in the second half. Assuming these players have the same average foul rate, the probability of Player A

⁵ The median minutes played per game for starters is approximately 30 minutes. Figure 2 includes just those players whose average minutes per game is below 30 minutes.

fouling out is obviously considerably lower than the probability of Player B fouling out, simply because he will be on the court for less time.

I also investigate whether better players are benched more or less often when they are in foul trouble. I use end-of-season VORP as a proxy for player quality and regress minutes played in each game on the number of minutes a player is in foul trouble, his VORP, and the interaction between the two (Table 1).

Table 1: Regression of minutes played on minutes in foul trouble and VORP

Constant	28.162*** (0.087)
Minutes in foul trouble	-0.324*** (0.019)
VORP	1.216*** (0.035)
VORP * Minutes in foul trouble	-0.017 (0.010)
R-squared	0.134
No. observations	12,300

Standard errors are reported in parentheses. *, **, *** indicates significance at the 90%, 95%, and 99% level, respectively.

First, the results of this regression again show that the more minutes a player is in foul trouble, the fewer minutes he plays in the game. In addition, the positive coefficient for VORP indicates that better players tend to play more minutes. However, the interaction term is not statistically significant. That is, the effect of foul trouble on playing time does not seem to be affected by the quality of the player.

While player quality overall does not significantly affect coaches' decisions, it is possible that star players are treated differently from non-star players. I identify a star player as any player who was named an all-star in the current season, or any of the two previous seasons (see Appendix 1 for full list). I regress minutes played in each game on the number of minutes a

player is in foul trouble, the dummy variable for whether or not the player is a star, and the interaction between the two. I also include VORP, in order to detect whether the effect of being a star player exists above and beyond just being a good player (Table 2).

Table 2: Regression of minutes played on minutes in foul trouble, star, and VORP

Constant	28.184*** (0.085)
Minutes in foul trouble	-0.354*** (0.016)
Star	0.829*** (0.182)
Star * Minutes in foul trouble	0.091** (0.043)
VORP	1.083 *** (0.040)
R-squared	0.136
No. observations	12,300

Standard errors are reported in parentheses. *, **, *** indicates significance at the 90%, 95%, and 99% level, respectively.

Indeed, there is a significant positive effect of being a star player, even controlling for players' VORP. Furthermore, star players are treated differently when in foul trouble. The positive coefficient on the interaction term in the model indicates that the time a player is in foul trouble has a smaller negative effect on playing time for star players as compared to non-star players.

The interaction term for VORP and minutes in foul trouble (-0.017) is non-significant, but if anything, it appears to exacerbate the effect of foul trouble (Table 1). The interaction term for star and minutes in foul trouble (0.091) diminishes the effect of foul trouble (Table 2). Given this apparent contradiction, it seems that the effect of foul trouble must vary non-linearly based on player quality. In order to investigate this further, I regressed minutes played on minutes in foul trouble for players who fall into five different buckets of VORP (see Table 3).

Table 3: Regression of minutes played on minutes in foul trouble, bucked by VORP

VORP Bucket	No. Observations	Coefficient
-2 to -.49	882	-0.224
-.5 to 0.99	4,082	-0.297
1 to 2.49	5,068	-0.418
2.5 to 3.99	1,250	-0.437
4 to 10	1,018	-0.206

These bucketed regressions indicate that foul trouble does indeed have the smallest negative effect on playing time for the best players, as one might expect based on the regression in Table 2. Interestingly, it appears that foul trouble has the biggest effect on playing time for players in the next tier—those who are above average but not elite.

One plausible reason coaches might bench their players who are in foul trouble is that they may believe those players are more likely to accumulate additional fouls. One can imagine why that might be the case. The fact that a player has already accumulated fouls might indicate that the way he is playing is making him more foul prone in that particular game. Or perhaps the offense might try to attack a player who is in foul trouble, making him more likely to get called for another foul. On the other hand, it seems equally likely that a player in foul trouble is less likely to be called for additional fouls. A player in foul trouble may intentionally play less aggressively. In addition, referees are less likely to call a foul on a player who is in foul trouble (Moskowitz & Wertheim, 2011). According to former referee Tim Donaghy⁶, this is not an accident:

“If Kobe Bryant had two fouls in the first or second quarter and went to the bench, one referee would tell the other two, ‘Kobe’s got two fouls. Let’s make sure that if we call a foul on him, it’s an

⁶ Tim Donaghy gained notoriety for betting on games that he officiated, and intentionally making calls to win those bets. Despite this obvious lack of integrity, there is no reason to think that his account of referees’ treatment of players in foul trouble is fabricated.

obvious foul, because otherwise he's gonna go back to the bench. If he is involved in a play where a foul is called, give the foul to another player” (Donaghy, 2010).

In order to address this question empirically, I first look at the foul rate for players who are not in foul trouble and the foul rate for players who are in foul trouble⁷. Overall, the average foul rate is 0.088 fouls per minute (4.23 fouls per 48 minutes). When not in foul trouble, players' average foul rate is 0.089 fouls per minute (4.26 fouls per 48 minutes). This is very similar to the overall foul rate, because players are generally not in foul trouble. When players are in foul trouble, however, the average foul rate is 0.055 fouls per minute (2.65 fouls per 48 minutes). Of the 112 players who logged at least ten minutes in foul trouble, 83 (74.11%) had a lower foul rate when in foul trouble than when not in foul trouble, $X^2 (1, N=112) = 25.08, p < .001$.

Because coaches treat star players differently when they are in foul trouble, one might wonder whether this might be related to how star players are refereed when they are in foul trouble. After all, when Tim Donaghy provided an example of a foul-ridden player on whom he did not want to call additional fouls, he used megastar Kobe Bryant. Unsurprisingly, star players are called for fewer fouls per minute ($M_{\text{star foul rate}} = 0.074$ fouls per minute) than are non-stars overall ($M_{\text{non-star foul rate}} = 0.091$ fouls per minute), $t(110) = 3.264, p = .001$. However, the effect of foul trouble does not seem to be different for stars and non-stars, as 75.00% of star players have a lower foul rate when in foul trouble and 73.91% of non-stars have a lower foul rate when in foul trouble, $X^2 (1, N=112) = 0.010, p = .920$. Taken together, this suggests that coaches are justified in allowing star players to play more often when in foul trouble, as they are generally less likely to be called for additional fouls. However, coaches would also benefit from

⁷ In this analysis, I only include the 112 players with more than 10 minutes played in foul trouble. The average foul rate is calculated as the average of players' individual rates, not as the total fouls for all players divided by the total minutes for all players.

recognizing that all players, not just stars, are less likely to be called for additional fouls when in foul trouble.

Discussion

Coaches consistently bench players who are in foul trouble, and appear to do so dogmatically using the Q+1 rule. These substitutions do not merely shift the minutes that players rest, but rather end up costing overall playing time. This seems to be true for all players, though to a lesser extent for star players. Furthermore, this behavior appears to be unnecessary, as players who are currently in foul trouble are significantly less likely to be called for additional fouls. In fact, in the 1,230 games played in the 2015-2016 regular season, only 144 starters fouled out (1.17%). Even if we look at just those players who were in foul trouble at any point in a game, just 6.13% of those players ended up fouling out.

Section 2: Does foul trouble affect performance?

Section 1 established that coaches are in fact sacrificing overall playing time by following the Q+1 rule. This might be a sensible strategy if having a player in the game in foul trouble is detrimental to the team's performance. In other words, it is possible that players in foul trouble do not play as well, and that this hurts the team. By benching the player when he is in foul trouble and bringing him back in the game later, a coach might be able to avoid this negative effect. Ultimately, the individual performance of the individual player in foul trouble is not necessarily what matters. I therefore begin by investigating this question on a team level, by examining how having a player on the court in foul trouble affects team performance. I then examine individual offensive and defensive performance for the players who are in foul trouble.

Data

For the primary analysis in this section, I use the raw play-by-play data to create a new data set such that each observation represents a “stint” in a particular game, similar to the structure of an adjusted plus-minus analysis (Rosenbaum, 2004; Jacobs, 2017). In this context, I define a stint as a stretch of time within a specific quarter in a game in which the same ten players are on the court with the same fouls status (i.e. in foul trouble vs. not in foul trouble). A stint therefore ends when a period ends, when there is a substitution, or when a player enters foul trouble. The resulting data set has 37,007 stints, with an average of 30.09 stints per game⁸. For each stint, I calculate the point differential per 100 possessions for each team.⁹ This “margin” is an estimate of what the expected point differential would be for an entire game if the single stint were extrapolated. So, for example, if a stint has eight total possessions (four for each team), and the home team wins the stint 5-4, the margin would be $(5/4 - 4/4) * 100 = 25$. In this data set, the average margin is comprised of 5.44 possessions.

In each stint, I also include the total difference in VORP between the players on the home team and players on the away team. This difference is meant to be a proxy for the difference in quality of the players on the court for the two teams. Alternatively, I could have included the total VORP of each team as separate variables. However because the dependent variable is point differential per possession, the difference in VORP is a more sensible variable. In other words,

⁸ Again, overtime periods are removed from this analysis.

⁹ The precise formula is $(\text{home points}/\text{home possessions} - \text{away points}/\text{away possessions}) * 100$. In stints where one team does not have any possessions, I use its end-of-season adjusted offensive rating, which is defined as “an estimate of points scored per 100 possessions adjusted for strength of opponent defense” (www.basketball-reference.com). Also note that if a free throw occurs after a substitution, the points and possession are attributed to the stint in which the foul occurred.

the VORP of one team is not particularly meaningful in isolation; it gains meaning only in its value relative to the other team's VORP.

Of course, the stints that are played with a player in foul trouble are not randomly assigned. Coaches presumably play their players in foul trouble when they believe it will be most efficacious. In what follows, I describe how the stints played with a player in foul trouble compare to the stints in which no foul-troubled players are in the game.

First, Table 4 provides the number of stints in the data set for each segment of the game, separated by whether or not a player on either team was in the game while in foul trouble during those stints. I define game segments as halves of quarters. So, for example, “1.1” represents the first half of the first quarter, and “3.2” represents the second half of the third quarter.

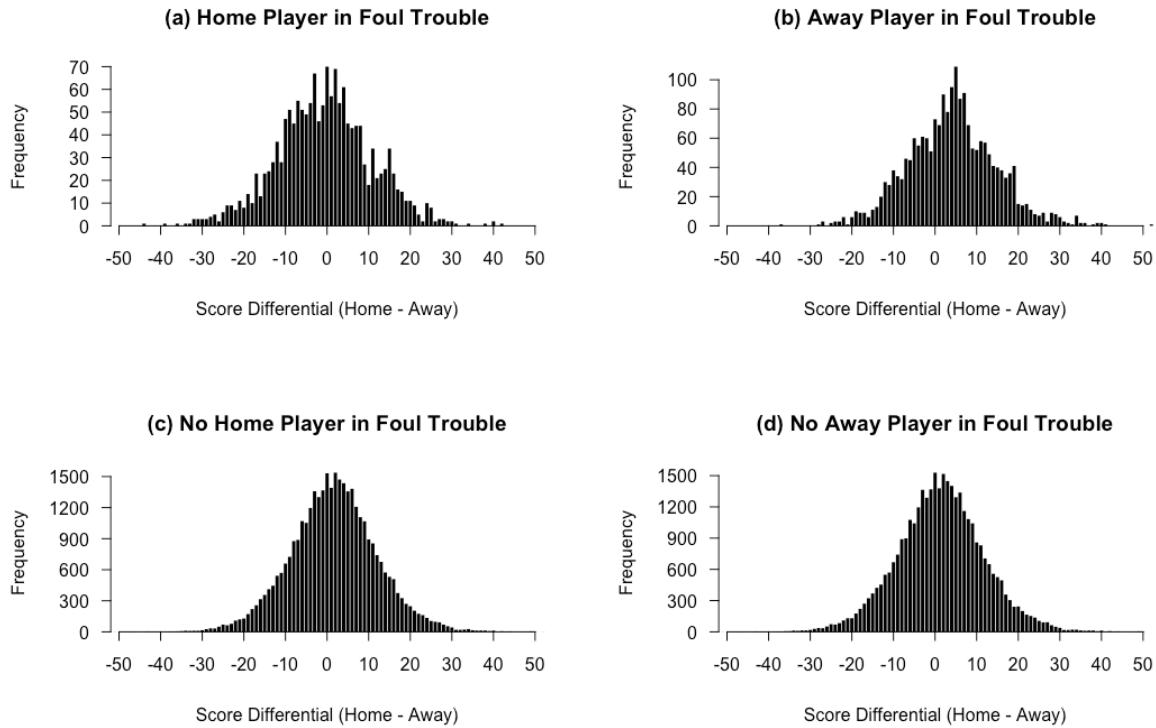
Table 4: Number of stints for each game segment, by foul trouble

	1.1	1.2	2.1	2.2	3.1	3.2	4.1	4.2	Total
Home only	34	307	20	101	61	311	58	541	1,433
Away only	28	365	25	155	69	425	72	733	1,872
Both	4	17	0	3	5	25	3	103	160
Neither	916	6,072	3,981	6,209	1,514	5,859	3,847	5,144	33,542
Total	982	6,761	4,026	6,468	1,649	6,620	3,980	6,521	37,007

There are two notable trends in this data. First, it is much more common for a player in foul trouble to be in the game in the latter half of each quarter. This is simply mechanical; the more time that has elapsed in a quarter, the more likely it is that a player will have been called for a threshold foul (i.e. a foul that moves a player who was not in foul trouble into foul trouble). Second, players are most likely to be in the game in foul trouble at the end of the fourth quarter. This is also predictable, as I suggest that coaches tend to bench players who are in foul trouble specifically so that they will be available for the final minutes of the game.

Next, I compare the distributions of score differentials (home score – away score) when players from each team are in the game in foul trouble versus not in foul trouble (Figure 3a-d).

Figure 3a-d: Distributions of score differential in each stint, by foul trouble

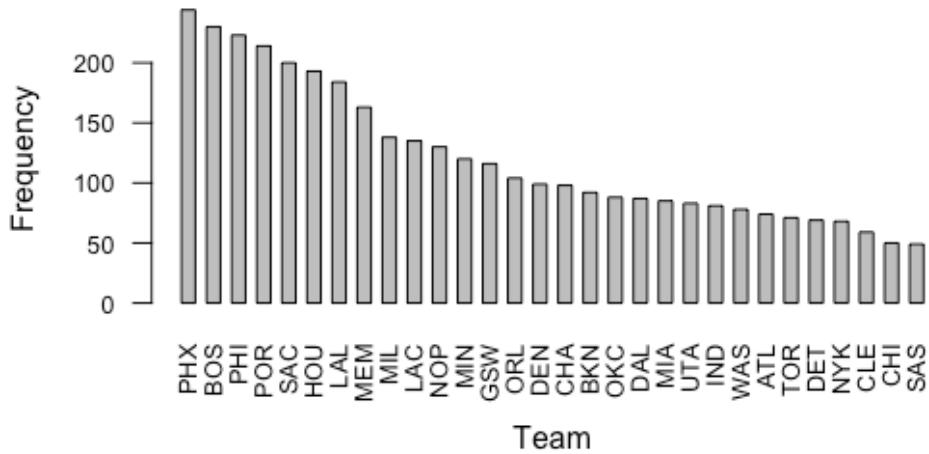


First, Figures 3a and 3b show that the score differentials when players in foul trouble are in the game are distributed fairly normally. However, these distributions are shifted compared to Figures 3c and 3d, such that teams are more likely to have a player in the game in foul trouble when they are losing than when they are winning. The average score differential when the home team has a player in the game in foul trouble ($M = -0.604$, $SD = 11.444$) is significantly lower than the average score differential when the home team does not have a player in foul trouble in the game ($M = 1.606$, $SD = 10.437$), $t(37,005) = 8.232$, $p < .001$. Similarly, the average score differential when the away team has a player in the game in foul trouble ($M = 4.101$, $SD = 10.619$) is significantly higher than the average score differential when the away team does not

have a player in foul trouble in the game ($M = 1.361$, $SD = 10.465$), $t(37,005) = 11.466$, $p < .001$ ¹⁰.

Finally, Figure 4 depicts the number of stints in the data set in which each NBA team had a player in the game in foul trouble.

Figure 4: Stints with a player in foul trouble, by team



There is considerable variation in how often each team's coach allows players in foul trouble to play. Of course, there are innumerable explanations for these differences. For example, this pattern might be attributable to coaches' philosophical differences. Or perhaps teams with fewer foul trouble stints tend to have better substitutes, allowing their coaches to feel more comfortable benching starters. It is also possible that this simply reflects the effects of score differential; perhaps some teams tend to be losing more often, and that is why they are more likely to play their players in foul trouble. In any case, this significant variation is notable.

There are presumably several other selection effects that I cannot account for in the present investigation. For instance, it is possible that some of these decisions are driven by specific player match-ups. As I detail below, in my analyses, I control for as many relevant

¹⁰ Score differential is calculated as home score – away score. Therefore, a higher point differential indicates that the away team is losing by more points.

factors as possible. Nonetheless, I acknowledge the possibility that other uncontrolled factors could be playing a role.

While I use the stint-by-stint data to evaluate the effect of foul trouble on team performance, I use players' individual offensive and defensive ratings (Oliver, 2004) in order to investigate the effect of foul trouble on individual performance. Offensive (and defensive) ratings are designed to estimate the points a player contributes (allows) per possession that he is responsible for. One can think of the construction of these statistics as a kind of accounting exercise, where each team point and possession is distributed to the players who are responsible for them. For example, if a player commits a turnover, this would negatively affect his offensive rating, as he is fully responsible for that possession but did not contribute any points. Similarly, if a player gets a steal, this positively affects his defensive rating, as he is fully responsible for that possession and did not allow any points. Traditional offensive and defensive ratings are calculated using end-of-game box score statistics, and so they necessarily rely on several estimations. For example, when looking at an end-of-game box score, there is no way to identify which defensive players were on the court when specific baskets were scored. As a result, a traditional defensive rating calculation weights the team's points allowed by the percent of time a given player was on the court.

In this analysis, however, I need to know precisely when specific possessions occur, in order to determine how players' foul status affects their performance. I therefore created a version of individual offensive and defensive ratings using play-by-play data, rather than end-of-game box scores (see Appendices 2 and 3 for further details). This technique also minimizes the need for estimators, and increases the accuracy of the ratings, particularly on offense (for details, see: Parker, 2009a and Parker 2009b). In order to compare these new ratings to their box-score

counterparts, I examined the correlations between each of these ratings and another commonly used metric – real plus-minus. Compared to the box-score offensive ratings, the new play-by-play offensive ratings are more highly correlated with offensive real plus-minus ($r_{\text{new off. rating}} = .652$ vs. $r_{\text{box off. rating}} = .555$). Relatedly, compared to the box-score defensive ratings, the new defensive ratings are slightly more highly correlated with defensive real plus-minus ($r_{\text{new def. rating}} = -.765$ vs. $r_{\text{box def. rating}} = -.761$)¹¹.

Nonetheless, offensive and defensive ratings are not perfect statistics. For example, some contend that offensive ratings overweight offensive rebounds, or that defensive ratings are too reliant on the defensive ability of a player's teammates (e.g. Johns, 2011; Larsen, 2013). In my analysis, though, I am comparing players' ratings to themselves at different times (i.e. in foul trouble vs. not in foul trouble). This makes these concerns much less problematic.

Analysis

Effect of foul trouble on team performance

In order to determine whether a player's foul trouble affects his team's performance, I created a model using the stint-by-stint data set. The primary model is given by the equation¹²:

$$\begin{aligned} \text{Margin} = & \alpha + \beta_1 \text{home.ft} + \beta_2 \text{away.ft} + \beta_3 \text{VORP.dif} + \beta_4 \text{VORP.dif.ft} + \beta_5 \text{score.dif} \\ & + \beta_6 \text{min.cat1.2} + \beta_7 \text{min.cat2.1} \dots \beta_{12} \text{min.cat4.2} \end{aligned}$$

where:

- Margin is calculated as (home points/home possessions – away points/away possession)*100,

¹¹ These correlations are negative, because better defense is associated with a lower defensive rating, but a higher defensive real plus-minus.

¹² In all regressions using the stint-by-stint data, I cluster standard errors by game.

- `home.ft` is a dummy variable indicating whether the home team has at least one player on the court in foul trouble,
- `away.ft` is a dummy variable indicating whether the away team has at least one player on the court in foul trouble,
- `VORP.dif` is the difference in VORP between the players on the home team and the players on the away team,
- `VORP.dif.ft` is the difference in VORP between the players in foul trouble in the game for the home team and the players in foul trouble in the game for the away team,
- `score.dif` is the current score differential in the game at the end of the stint,
- and `min.cat1.2`, `min.cat2.1`...`min.cat4.2` are fixed effects for the segment of the game in which the stint ended (e.g. `min.cat1.2` is the second half of the first quarter and `min.cat3.1` is the first half of the third quarter).

Remember that an increase in the margin indicates that the home team is playing better relative to the away team. Similarly, a decrease in the margin indicates that the away team is playing better relative to the home team. In order to further elucidate the structure of the model, I will walk through the logic of the coefficient for `VORP.dif`. One should anticipate that the coefficient for `VORP.dif` is positive. If the home team has better players on the court than the away team, `VORP.dif` will be positive, and on average, the margin should also be positive. Conversely, if the away team has better players on the court than the home team, `VORP.dif` will be negative, and on average, the margin should also be negative.

Therefore, if foul trouble hinders a team's performance, the coefficient on `home.ft` should be negative, while the coefficient on `away.ft` should be positive. That is, having a player on the court in foul trouble for the home team should decrease the margin, while having a player on the

court in foul trouble for the away team should increase the margin. In reality, I find the opposite result. It appears that having a player in the game in foul trouble is associated with an improvement in team performance (Table 5).

Table 5: Regression of scoring margin on each team's foul trouble, with controls

Constant	3.659 (2.039)
home.ft	6.194* (3.677)
away.ft	-8.257*** (3.307)
VORP.dif	0.681*** (0.088)
VORP.dif.ft	-1.382 (1.660)
score.dif	2.064*** (.053)
minute category (1.1 - 4.2)	
R-squared	0.068
No. observations	23,699

Standard errors are reported in parentheses. *, **, *** indicates significance at the 90%, 95%, and 99% level, respectively. Note: this model excludes any stint with fewer than four possessions (i.e. two possessions per team)

Robustness Checks

Of course, when building the primary model above, I made several decisions regarding which variables and observations to include or exclude. I therefore conducted a series of robustness checks. Below, I describe the various factors that I tested in these robustness checks, and the conclusions that I draw from them. The detailed results of each model are presented in Tables 10 and 11 in Appendix 4.

One relevant decision was whether or not to include control variables. Because my goal was to try to isolate the effect of foul trouble as much as possible, and thereby minimize possible

selection effects, I decided to include the control variables described above. I included VORP.dif to account for the possibility that the effect of foul trouble might be related to the quality of players on the court for each team. This also indirectly addresses potential team differences, though I include team fixed effects as a robustness check (Appendix 4, Table 10, Column 10). I included VORP.dif.ft because I wanted to understand the effect of any player being in foul trouble, independent from the quality of that player. Score.dif was included in the model because as I described previously, coaches are more likely to allow foul-troubled players to play when the team is losing. Similarly, I included the minute category to address the fact that players are more likely to be in the game in foul trouble in some game segments than in others. Indeed, when I run the model without these control variables, the effect of foul trouble does not emerge (Appendix 4, Table 10, Column 1).

Next, I decided to include only those stints with at least four possessions. The more possessions there are in a stint, the less noise there is in that stint's data. In particular, stints in which one team does not have any possessions are especially noisy, because that team's points-per-possession is estimated based on its end-of-season adjusted offensive rating. Nonetheless, it is possible that this restriction might have had unintended effects. I therefore tested various models with and without this restriction. Overall, including all observations slightly reduces the observed effects of foul trouble, but does not change the pattern of results nor the ultimate conclusion that foul trouble is associated with better team performance (Appendix 4, Table 10, Columns 2, 4, and 6).

There were several different ways in which I could have operationalized foul trouble. I ultimately decided to include a single dummy variable for each team indicating whether or not at least one player on that team was in the game in foul trouble. However, in my robustness checks,

I include two other operationalizations. First, I include a dummy variable for each team indicating whether one and only one player on that team is in the game in foul trouble. In this version, I exclude from the analysis any stint in which more than one player on a team is in the game in foul trouble (Appendix 4, Table 10, Columns 4 and 5). Second, I include two dummy variables for each team indicating (1) whether one and only one player is in the game in foul trouble, and (2) whether two or more players are in the game in foul trouble (Appendix 4, Table 10, Columns 6 and 7). Regardless of how I classify foul trouble, the pattern of results consistently indicate that, if anything, foul trouble is associated with better team performance.

Next, some stints obviously have an odd number of total possessions, which means that one team has more possessions in that stint than does the other team. I attempt to address this with my dependent variable, by calculating the margin as the difference in home team points per home team possession and away team points per away team possession. Nonetheless, it is possible that stints with an odd number of possessions might affect the model's results. For instance, it is possible that in stints in which a player is in the game in foul trouble, that player's team is more likely to start on offense, and so that team might be more likely to have an additional possession. While I believe the structure of the dependent variable minimizes this concern, I ran a robustness check that includes just those stints with an even number of possessions. The data is noisier due to the reduced sample size, but the pattern of results is the same (Appendix 4, Table 10, Column 8).

In Section 1, I found that coaches treat star players differently from non-star players when they are in foul trouble, as their overall playing time is less affected by the time they are in foul trouble. I therefore ran a model that includes dummy variables indicating whether a star player was in the game while in foul trouble for each team. While the coefficients are not

significant predictors of the scoring margin, directionally, they indicate that stars in foul trouble might improve team performance even more than non-stars (Appendix 4, Table 10, Column 9).

Next, as previously mentioned, coaches are most likely to play their players in foul trouble at the end of the game. While I control for game segment in my primary model, I also ran the model independently on different segments of the game (i.e. first three quarters only versus fourth quarter only). The precise coefficient estimates vary, and interestingly, the away team's foul trouble seems to have a larger effect earlier in the game, while the home team's foul trouble seems to have a larger effect later in the game. Nonetheless, the pattern of results consistently indicates that if anything, foul trouble is associated with improved performance (Appendix 4, Table 11).

And finally, the distribution of the margin is non-normal, and includes several outliers. In order to address this concern, I ran a quantile regression using median values. Again, the pattern of results is unchanged (Appendix 4, Table 10, Column 11).

What contributes to the improvement?

The positive effect of foul trouble on performance is somewhat counterintuitive, and therefore necessitates further investigation. It appears that there are two primary factors contributing to this difference, namely the rate at which teams are called for additional fouls and teams' field goal percentages¹³. First, using the same model structure as my primary model above, I replace the scoring margin with the away team's foul rate (i.e. fouls per possession), in order to evaluate the association between each team's foul trouble and the rate at which the away team is called for additional fouls. I run similar models on the home team's foul rate, the away

¹³ Field goal percentage is calculated as (two-point shots made + three-point shots made)/(two-point shots attempted + three-point shots attempted).

team's field goal percentage, and the home team's field goal percentage. The results of these four models are presented in Table 6.

Table 6: Regressions of team foul rates and team field goal percentages

	Home foul rate	Away foul rate	Home field goal %	Away field goal %
Constant	0.173*** (0.006)	0.199*** (0.007)	0.527*** (0.008)	0.529*** (0.008)
home.ft	-0.004 (0.010)	0.047*** (0.011)	-0.003 (0.013)	-0.027** (0.014)
away.ft	0.031*** (0.008)	-0.032*** (0.008)	-0.016 (0.012)	0.017 (0.012)
VORP.dif	-0.002*** (0.000)	0.016*** (0.000)	0.001*** (0.000)	-0.001*** (0.000)
VORP.dif.ft	0.003 (0.004)	-0.007 (0.005)	-0.005 (0.006)	0.007 (0.006)
score.dif	0.000 (.000)	0.000 (.000)	0.005*** (0.000)	-0.005*** (0.000)
minute category				
No. observations	23,699	23,699	23,191 ^a	23,167 ^b

Standard errors are reported in parentheses. *, **, *** indicates significance at the 90%, 95%, and 99% level, respectively. Note: these models exclude any stint with fewer than four possessions (i.e. two possessions per team)

a: This only includes stints in which the home team attempted at least one field goal.

b: This only includes stints in which the away team attempted at least one field goal.

Table 6 indicates that when the home team has a player in the game in foul trouble, the away team is significantly more likely to be called for additional fouls, and the away team's field goal percentage is significantly worse. When the away team has a player in the game in foul trouble, the away team is significantly less likely to be called for additional fouls, and the home team is significantly more likely to be called for additional fouls.

Next, I include these four variables (home team foul rate, away team foul rate, home team field goal percentage, and away team field goal percentage) as covariates in the model predicting scoring margin. If the positive association between foul trouble and team performance is driven

by changes in foul rates and field goal percentages, then including these variables in the model should eliminate the observed effects of foul trouble. The results of Table 7 demonstrate that this is precisely the case.

Table 7: Regression of scoring margin controlling for foul rates and field goal percentages

Constant	4.530*** (1.498)
home.ft	0.682 (2.076)
away.ft	1.500 (1.921)
VORP.dif	0.149*** (0.057)
VORP.dif.ft	0.647 (0.841)
score.dif	0.542*** (.033)
h.foul.rate	-45.645*** (1.791)
a.foul.rate	40.336*** (1.737)
h.fgp	154.969*** (1.151)
a.fgp	-156.293*** (1.194)
minute category (1.1 - 4.2)	
R-squared	0.727
No. observations	22,688 ^a

Standard errors are reported in parentheses. *, **, *** indicates significance at the 90%, 95%, and 99% level, respectively. Note: this model excludes any stint with fewer than four possessions (i.e. two possessions per team)
^a: This only includes stints in which both the home team and away team attempted at least one field goal.

There is a somewhat intuitive explanation for the observed relationship between foul trouble and foul rates. Referees are less likely to call fouls on teams with foul-troubled players in the game. And on plays when there is noticeable contact between two players and a foul needs to

be called on someone, it makes sense that the opposing team is more likely to be called for those fouls. One might wonder, though, whether this is really a result of foul trouble, or if instead, referees are merely trying to be as balanced as possible and teams with a player in foul trouble are more likely to have been called for more fouls overall. It turns out that the total number of fouls on each team does predict each team's foul rate, just as foul trouble does (Home foul rate: $b_{\text{home total fouls}} = -0.001$, $t(23,686) = -1.208$, $p = .227$; $b_{\text{away total fouls}} = 0.004$, $t(23,686) = 6.901$, $p < .001$. Away foul rate: $b_{\text{home total fouls}} = 0.004$, $t(23,686) = 6.939$, $p < .001$; $b_{\text{away total fouls}} = -0.002$, $t(23,686) = -2.490$, $p = .013$). However, even controlling for total fouls on each team, having a player in the game in foul trouble still significantly predicts foul rates (see Table 8).

Table 8: Regressions of foul rates, controlling for total fouls

	Home foul rate	Away foul rate
Constant	0.173*** (0.006)	0.199*** (0.007)
home.ft	-0.005 (0.010)	0.035*** (0.011)
away.ft	0.020** (0.009)	-0.033*** (0.008)
home.total.fouls	-0.001 (0.001)	0.004*** (0.001)
away.total.fouls	0.004*** (0.001)	0.001* (0.001)
VORP.dif	-0.002*** (0.000)	0.002*** (0.000)
VORP.dif.ft	0.003 (0.004)	-0.007 (0.005)
score.dif	-0.000 (0.000)	0.000 (0.000)
minute category (1.1 - 4.2)		
R-squared		0.233
No. observations		23,699

Standard errors are reported in parentheses. *, **, *** indicates significance at the 90%, 95%, and 99% level, respectively. Note: this model excludes any stint with fewer than four possessions (i.e. two possessions per team)

Based on these analyses, it appears that referees may be sensitive to multiple signals in their efforts to maintain balance in their foul calls. One of these signals seems to be whether a player on the floor is in foul trouble. In addition to the way referees are calling fouls, players in foul trouble also may be playing more tentatively. The relationship between foul trouble and field goal percentage, though, is more difficult to explain. Perhaps future research might help shed light on this counter-intuitive result.

Effect of foul trouble on individual performance

The primary model (see Table 5) indicates that teams tend to play better when they have a player on the court in foul trouble. But how does foul trouble affect the player's individual performance? To address this question, I first compare players' offensive ratings when they are not in foul trouble to their offensive ratings when they are in foul trouble. Because many players do not have extensive minutes played in foul trouble, their offensive ratings are fairly noisy. I will therefore analyze the proportion of players whose ratings improve vs. worsen when in foul trouble. I find that out of 358 players who logged at least one offensive possession while in foul trouble, 231 (64.53%) have a lower offensive rating when they are in foul trouble than when they are not, $\chi^2(1, N = 358) = 29.634$, $p < .001$.

Players' defensive ratings do not seem to be affected by foul trouble. Of the 375 players who log at least one defensive possession in foul trouble, 194 (51.73%) have a lower defensive rating when in foul trouble, $\chi^2(1, N = 375) = 0.384$, $p = .536$. It is important to note, however, that individual defensive ratings, even when calculated using play-by-play data, are quite noisy. For instance, the raw data does not provide information regarding which player was guarding the

shooter on any given shot, and so much of a player's individual defensive rating depends on his team's performance.

Discussion

The starting premise of this part of the analysis was that if a team does not play significantly worse when a player with foul trouble remains in the game, then that player should probably not be benched. Based on these findings, it appears that players in foul trouble do not play as well individually, but their presence on the court significantly improves their team's performance. Given that the differences in team performance seem to be driven by foul rates, it is perhaps easier to reconcile this apparent contradiction. While it is certainly counter-intuitive that having a foul-troubled player in the game improves team performance, Moskowitz and Wertheim (2011) and Maymin, Maymin, and Shen (2012) allude to similar findings. To be clear, I am certainly not suggesting that players should intentionally get themselves into foul trouble. Nonetheless, these findings demonstrate that it is quite costly for coaches not to play players who are in foul trouble. In so doing, a coach is not only substituting the value of a lesser player for the value of a better player, but he is also missing out on the additional boost that a team seems to get by having a player in foul trouble in the game.

Section 3: Is the end of the game meaningfully different from the rest of the game?

Thus far, I have established that coaches sacrifice overall playing time by benching players who are in foul trouble, and that team performance is actually improved by having a player in foul trouble in the game. The only way this strategy might be optimal, then, is if the end of the game is somehow different from the rest of the game. I will address three potential ways in which the end of the game might be unique. First, it is possible that the pace accelerates towards

the end of the game, such that there are more possessions per minute at the end of a game than there are earlier in the game. If this were true, coaches might be justified in saving their best players for late in the game, as having them in the game for more possessions is probably more meaningful than having them in the game for more minutes.

Second, some might contend that players try harder at the end of the game, so having the team's best players available then is most important. This implies that there should be a stronger relationship between skill level and outcome at the end of the game. In other words, point differential should be more strongly predicted by the difference in skill between the players on the court for each team at the end of the game than at any other time.

And finally, it is possible that some kinds of players are more valuable at the end of the game than they are the rest of the game. For example, qualitatively, it seems that there is more one-on-one play and more free throws at the ends of close games. Maybe that makes certain kinds of players more or less valuable. As a proxy, I address this question by comparing the difference in star players' performance from the beginning of the game to the end of the game, to the difference in non-star players' performance from the beginning of the game to the end of the game.

Data

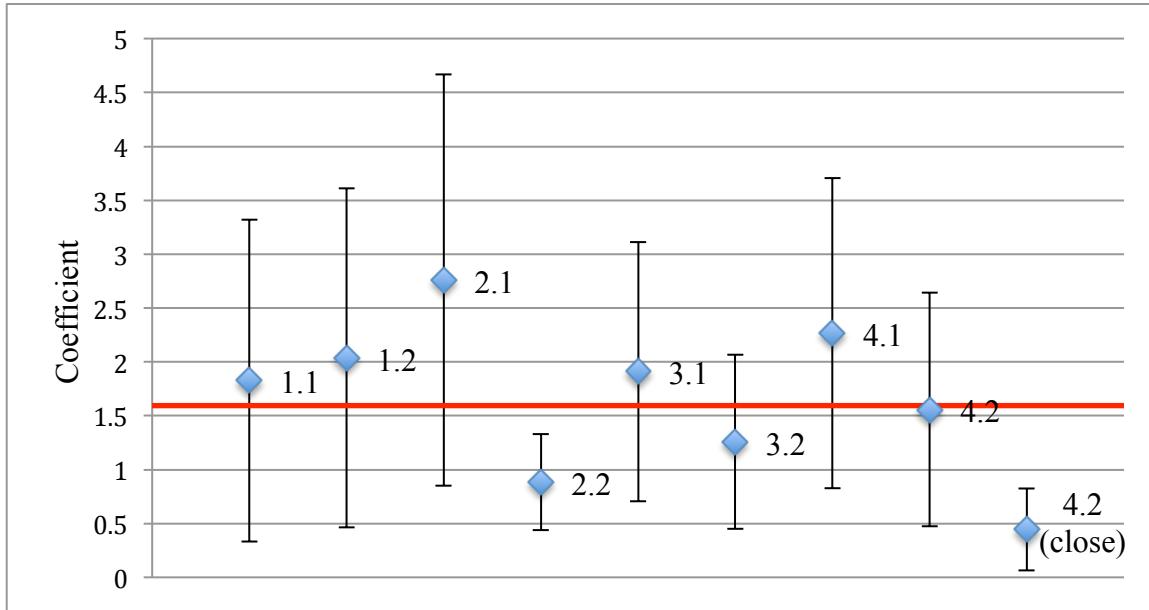
First, in order to determine the pace for different segments of the game, I used the raw play-by-play data to calculate the number of possessions per 48 minutes for the segment of interest. The rest of the data sets used in this section are very similar to those used in Section 2. I use a similar stint-by-stint dataset, except that for this analysis, a new stint does not begin when a player enters foul trouble. This section also uses offensive and defensive ratings, which were described in Section 2.

Analysis

Pace is typically measured as possessions per team per 48 minutes. A paired t-test reveals that the pace in the first three quarters of games ($M = 95.57$, $SD = 5.26$) is faster than in the fourth quarter of games ($M = 93.71$, $SD = 8.04$), $t(1229) = 7.924$, $p < .001$. However, perhaps close games are different. Is the pace faster in the fourth quarter than in the first three quarters if we look only at games that are close to start the fourth quarter? I first look at games in which the score differential is in single digits (i.e. less than a 10-point difference). Again, a paired t-test reveals that the pace is slower in the fourth quarter of these games ($M = 93.40$, $SD = 8.24$) than in the first three quarters ($M = 95.13$, $SD = 5.18$), $t(674) = 5.186$, $p < .001$. Looking at even closer games (games within five points going into the fourth quarter), a paired t-test reveals that the pace is still slower in the fourth quarter ($M = 93.40$, $SD = 8.24$) than in the first three quarters ($M = 94.91$, $SD = 4.87$), $t(406) = 4.439$, $p < .001$.

Next I examine whether the skill level of the players on the court for each team is more predictive of outcomes towards the end of the game. Each point in Figure 5 represents a regression of scoring margin on the difference in VORP for a unique segment of the game (full regression output available in Appendix 5). Looking at the regressions separated by segment, it appears that skill difference is no more predictive at the end of the fourth quarter than at other times in the game. And in fact, difference in VORP is considerably less predictive of outcomes in the final minutes of close games (defined as a score differential within five points). I find a similar of results using other individual player metrics (see Appendix 5).

Figure 5: Coefficient for VORP.dif on scoring margin, by game segment



Nonetheless, it is possible that certain kinds of players are more valuable at the ends of games. In this analysis, I examine how star players' and non-star players' offensive and defensive ratings change over the course of the game. Star players are significantly more likely to have a higher offensive rating in the fourth quarter than the first three quarters (74.36%) than are non-star players (55.63%), $X^2(1, N = 465) = 5.117, p = .024$. Star players are also significantly more likely to have a better defensive rating in the fourth quarter than in the first three quarters (64.10%) than are non-star players (47.58%), $X^2(1, N = 472) = 3.911, p = .048$. This indicates that stars indeed appear to be more valuable at the ends of games.

Discussion

Overall, the difference in skill level of a team's players is less predictive of outcomes at the ends of close games as compared to the rest of the game. However, star players seem to be more effective at the ends of games, as compared to non-star players. This latter point might make it seem sensible to bench star players when in foul trouble in order to ensure that they are

available for the end of the game, which based on my analysis in Section 1, coaches indeed do. However, first, the drop-off in performance between a star player and his substitute tends to be fairly large. If that difference is bigger than the difference between a star player early in the game and a star player later in the game, then a coach should never sacrifice playing time to ensure a star's availability in the fourth quarter. Second, part of a star player's improved performance in the fourth quarter is likely due to opportunity. For example, stars may be more likely to take additional free throws towards the end of the game. If a star has fouled out and is no longer available, those opportunities do not disappear, but rather are distributed to the other players on the court.

General Discussion

At a baseline level, a coach gives his team the best chance of winning by playing his best players as many minutes as possible (while obviously giving them adequate time to rest). My analysis shows that players who are in foul trouble end up playing significantly fewer minutes, despite the fact that players in foul trouble are even less likely to be called for additional fouls. This strategy might make sense if the team's performance were hindered by a player's foul trouble. But in fact, teams perform even better when a player who is in foul trouble is in the game. And finally, a coach's decision to save a player in foul trouble for the end of the game might be justifiable if the end of the game were substantively different from the rest of the game. Not only is skill level not more predictive of outcomes at the ends of close games, but it is actually less predictive. Taken together, this analysis demonstrates that in general, coaches should not bench their players because of foul trouble.

Theoretical implications

As previously mentioned, this adds to a growing list of sports situations in which both coaches and fans make a systematic error in judgment. In order to better understand if and why people believe that players in foul trouble should be benched, I conducted a study with 400 NBA fans using Amazon Mechanical Turk. All participants read the following scenario:

Two teams, Team A and Team B, are locked in an evenly matched game. There are now 5 minutes left in the first half. One of Team A's best players, Bill Johnson, is called for his 3rd foul. The coach of Team A, recognizing that Johnson is in foul trouble, is deciding whether or not to take Johnson out of the game.

Approximately half of the participants were then told that the coach's only goal was to win the game, were required to confirm that they understood the coach's only goal was to win the game, and were then asked whether they thought the coach should bench Johnson or keep him in the game. Those who said that he should be benched were then asked why they held that belief. Specifically, they were asked to choose between the following options:

- The coach should save Johnson for later in the game
- Johnson will not play as well when he is in foul trouble
- If Johnson were to foul out, that would be bad for team morale
- Other

Of the 102 fans who said Johnson should be benched, 78 (76.47%) selected "the coach should save Johnson for later in the game" as their explanation.

The other half of participants, rather than being told that the coach's only goal was to win the game, were told that the coach's only goal was to maximize Johnson's playing time. They, too, were required to confirm that they understood that that was the coach's only goal and were then asked whether the coach should bench Johnson or keep him in the game. Of course, in reality the logically correct answer is that the coach should keep Johnson in the game. Nonetheless, 92 out of 202 people (45.54%) said that the coach should bench the player.

These results not only suggest a critical misunderstanding, but also underscore the overwhelming intuition that a player who is in foul trouble will end up fouling out. Fans who say that the coach should bench Johnson to save him for later in the game, and fans who believe that if Johnson stays in the game in foul trouble he will end up playing fewer minutes than if he is benched, presumably hold those beliefs because they are assuming that Johnson will foul out if he continues playing. Based on my analyses of average foul rates, the decreased foul rates of players who are in foul trouble, and the very low rate of foul-outs in the NBA, this intuitive belief is clearly unfounded.

This work also challenges the efficacy of using models of win probability for certain kinds of decisions. Win probability is a great tool for decisions that involve a single point in time. For example, consider a baseball game in which the game is tied in the bottom of the ninth inning, and the batting team has a runner on first base with no outs. A savvy manager might be interested in knowing the probability of winning the game with a runner on first base and zero outs (the current situation) versus the probability of winning the game with a runner on second base and one out (the likely resulting situation if he chooses to sacrifice bunt). However, win probability becomes problematic when they are used to make decisions over time. Win probabilities are more volatile towards the ends of games, as a single possession in the closing seconds can shift a team's chances of winning from a coin flip to a nearly certain victory. For this reason, models of win probability tend to underweight early events that end up being critical.

The same reasons that make a win probability model problematic for decisions over time can also help shed light on people's intuition that the end of the game is more important than the rest of the game. Win probabilities can shift more at the ends of close games, and simultaneously feel more important, because in the closing minutes of close games, it is known that every point

is critical. Conversely, the impact of points scored at the beginning of the game are unknowable at the time they are scored, and their impact is only appreciable in retrospect. However, when making a decision at the beginning of the game that will affect the end of the game (i.e. benching a player now to save him for later), the importance of future points are just as uncertain as the importance of early points. That is, if the game ends up being a blowout, then the end of the game will not matter at all, while if it ends up being close, the end of the game will obviously be very important. Ultimately, the scoreboard is indifferent to the timing of made baskets.

The cost of following the Q+1 strategy

In this section, I provide an estimated cost, both in terms of points and wins, that each team incurred by benching players who were in foul trouble. The estimates of each step of the process described below can be found in Table 9. First, for each team, I estimated the expected decrease in playing time associated with every minute a player is in foul trouble. Using player-games as observations (82 games * 5 starters per game = 410 observations per team), I regressed minutes played on total minutes in foul trouble, with fixed effects for players. This is similar to the regression model used in Section 1. For the Washington Wizards, for example, every minute a player was in foul trouble was associated with 0.478 fewer minutes played in the game (Table 9, “Coef. on ft. min.”). I then estimated the total time missed by starters due to foul trouble by multiplying that coefficient by the total number of minutes each team’s starters were in foul trouble over the course of the season. The Wizards’ starters were in foul trouble for a total of 564.950 minutes, and so I estimate that their starters missed a total of 270.083 minutes due to foul trouble ($0.478 * 564.950$) (Table 9, “Est. min. missed”).

That, however, does not account for foul outs. I therefore simulated what would happen if coaches left players in the game when they were in foul trouble, using the average foul rates for

players not in foul trouble and the average foul rates for players who are in foul trouble. The full details of this simulation can be found in Appendix 6. This simulation allowed me to calculate the expected minutes missed due to foul outs for every minute players are in foul trouble. On average, players would miss 0.085 minutes for every minute they were in foul trouble. I estimated the total time over the course of the season that players would have missed due to foul outs by multiplying 0.085 by the total number of minutes players were in foul trouble over the course of the season. Note that this assumes the same average foul rates apply to all players and all teams. So, for the Wizards, their starters would have missed approximately 47.809 minutes (0.085×564.950) if they were never benched because of foul trouble (Table 9, “Sim. min. missed”).

The cost of coaches’ actual substitution patterns then, in terms of minutes, can be calculated by subtracting the simulated time players would have missed due to foul outs from the actual time players missed due to foul trouble. The Wizards actually missed an estimated 270.083 minutes due to foul trouble. They would have missed an estimated 47.809 minutes due to foul outs if they were not benched when in foul trouble. Therefore, the cost of benching for the Wizards, in terms of minutes, is $270.083 - 47.809 = 222.274$ minutes (Table 9, “Min. missed Q+1”). In order to translate minutes into possessions, I simply multiplied the number of minutes missed for each team by that team’s average pace divided by 48 (pace is defined as possessions per team per 48 minutes). The Wizards’ average pace was 100.63, so their starters missed an estimated 465.988 possessions ($222.274 \times 100.63 / 48$) (Table 9, “Pos. missed Q+1”).

The next step was to estimate the number of points each team cost themselves by benching their starters. I calculated the average VORP of each team’s five most common starters, weighted by their average number of minutes played per game (Table 9, “Start. VORP”). I

similarly calculated the average VORP of each team's backups (any player who is not one of the five most common starters) weighted by their average number of minutes played per game (Table 9, "Back. VORP"). The difference between these two values estimates the average cost, per 100 possessions per team, of having a backup in the game instead of a starter. The weighted averages of the Wizards' starters' VORP and backups' VORP were 2.085 and 0.355, respectively (difference of 1.730) (Table 9, "VORP dif."). In addition, in Section 2 I demonstrated that teams tend to play even better with a foul-troubled player in the game. I therefore added 7.226 points per 100 possessions (the average of the effect for home team and away team) to the difference between starters and backups. By multiplying that value by the number of possessions that starters missed due to foul trouble substitutions for each team (and dividing by 100), I calculated an estimate of the points each team lost by benching players in foul trouble. For every 100 possessions that the Wizards' starters missed due to foul trouble, they cost themselves approximately 8.956 points ($1.730 + 7.226$). Their starters missed an estimated total of 465.988 possessions due to the use of the Q+1 strategy. Therefore, as a team, they missed out on a possible 41.736 points ($8.956 * 465.988/100$) (Table 9, "Est. points missed").

Finally, I estimated the number of wins those lost points might be worth, using the "Modified Pythagorean Theorem" (Morey, 1993). Originally developed by Bill James for use in baseball (James, 1983), the Pythagorean formula estimates the number of games a team "should" have won based on the points that team scored and the points they allowed over the course of a season. In basketball, the formula is:

$$Wins = \frac{points^{13.91}}{points^{13.91} + opponent\ points^{13.91}}$$

I assume that the points a team missed out on by following the Q+1 strategy would be evenly distributed between points scored and points allowed. The Wizards actually scored 8,534 points over the course of the season (Table 9, “Team pts.”). I estimate that if they did not bench players in foul trouble, they would have scored 8,554.868 points ($8,534 + 41.736/2$) (Table 9, “Adj. team pts.”). Based on these estimates, the Washington Wizards incurred the largest cost by benching players in foul trouble. Based on their actual points scored and points allowed, they were

expected to win 39.634 games ($\frac{8,534^{13.91}}{8,534^{13.91} + 8,575^{13.91}} = 39.634$) (Table 9, “Exp. wins”). Based on

their adjusted points scored and points allowed, they could have won 41.025 games

($\frac{8,554.868^{13.91}}{8,554.868^{13.91} + 8,554.132^{13.91}} = 41.025$) (Table 9, “Adj. exp. wins”). Therefore, they could have

won an estimated additional 1.359 games by not benching their players in foul trouble ($41.025 - 39.634 = 1.359$) (Table 9, “Wins missed”). On average, teams could have won an additional

0.634 games.

Table 9: Cost of benching players in foul trouble

Team	Coef. on ft. min. in missed	Total min. in missed	Est. min. in missed	Sim. min. in missed	Min. miss. Q+1	Pace	Pos. miss. Q+1	Start. VORP	Back. VORP	VORP dif.	Est. points missed	Team pts. pts.	Opp. pts. pts.	Adj. team pts.	Adj. opp. pts.	Exp. wins	Adj. exp. wins	Wins missed
WAS	0.478	564.950	270.083	47.809	222.274	100.63	465.988	2.085	0.355	1.730	41.736	8534	8575	8554.868	8554.132	39.634	41.025	1.391
DET	0.480	503.967	242.016	42.649	199.368	97.39	404.509	1.799	0.010	1.789	36.468	8361	8311	8379.234	8292.766	42.709	43.953	1.243
MIL	0.356	704.800	251.079	59.644	191.435	96.55	385.063	1.632	-0.022	1.653	34.190	8122	8465	8139.095	8447.905	29.520	30.612	1.092
IND	0.321	711.400	228.117	60.203	167.914	98.99	246.288	2.344	0.449	1.894	31.583	8377	8237	8392.791	8221.209	45.784	46.850	1.066
DEN	0.471	515.933	242.984	43.661	199.322	98.16	407.614	1.093	0.258	0.834	32.855	8355	8609	8371.428	8592.572	32.582	33.645	1.064
UTA	0.398	482.583	192.023	40.839	151.184	93.26	293.739	2.184	-0.001	2.185	27.644	8010	7863	8023.822	7849.178	46.253	47.227	0.974
NYK	0.461	459.933	211.971	38.922	173.049	95.83	345.485	1.372	0.231	1.141	28.906	8065	8289	8079.453	8274.547	33.281	34.258	0.977
HOU	0.238	826.567	190.022	70.795	128.227	100.08	267.354	3.118	-0.447	3.565	28.850	8737	8721	8751.425	8706.575	41.523	42.465	0.942
OKC	0.312	646.633	202.022	54.722	147.300	99.37	304.942	4.263	-0.023	4.286	35.105	9038	8441	9055.552	8423.448	59.141	60.052	0.911
MIN	0.351	535.933	188.167	45.354	142.813	97.59	290.356	1.606	-0.112	1.718	25.970	8398	8688	8410.985	8675.015	31.495	32.320	0.824
ORL	0.311	566.867	185.509	50.510	134.998	98.16	276.071	1.679	0.318	1.361	23.707	8369	8502	8380.853	8490.147	36.522	37.315	0.793
NOP	0.348	610.833	212.393	51.692	160.701	98.91	331.144	0.771	0.388	0.383	25.198	8423	8734	8435.599	8721.401	30.875	31.665	0.791
PHX	0.251	1066.083	267.235	90.218	177.017	100.86	371.957	0.426	0.154	0.272	27.891	8271	8817	8284.946	8803.054	23.884	24.661	0.777
SAC	0.208	925.500	192.941	78.321	114.620	102.24	244.141	1.385	0.272	1.113	20.360	8740	8943	8750.180	8932.820	34.508	35.149	0.642
TOR	0.302	473.800	143.039	40.096	102.943	95.31	204.407	2.796	0.807	1.989	18.835	8422	8053	8431.418	8043.582	53.378	53.968	0.590
PHI	0.247	967.867	239.186	81.907	157.279	100.23	328.418	0.414	-0.339	0.753	26.205	7988	8827	8001.102	8813.898	16.361	16.938	0.577
BKN	0.337	451.083	152.188	38.173	114.014	97.37	231.283	0.608	0.604	0.004	16.722	8669	8406	8677.361	8397.639	49.653	50.185	0.532
LAL	0.223	728.500	162.603	61.650	100.953	97.99	206.090	-0.057	0.022	-0.079	14.730	7982	8766	7989.365	8758.635	17.517	17.857	0.340
CHI	0.293	399.767	117.096	33.831	83.265	98.25	170.433	2.024	0.263	1.761	15.316	8335	8456	8342.658	8448.342	36.904	37.420	0.516
CLE	0.287	357.800	102.835	30.279	72.556	95.48	144.326	3.337	0.291	3.046	14.825	8555	8063	8562.413	8055.587	56.995	57.425	0.430
MIA	0.247	411.867	101.614	34.855	66.759	95.75	133.171	1.942	0.130	1.812	12.036	8204	8069	8210.018	8062.982	45.710	46.126	0.416
ATL	0.310	281.083	87.236	23.787	63.449	99.40	131.392	2.649	0.446	2.202	12.388	8433	8137	8439.194	8130.806	50.984	51.384	0.400
BOS	0.198	576.633	114.071	48.798	65.273	101.15	137.550	2.299	-0.091	2.390	13.226	8089	8692	8095.613	8685.387	22.052	22.407	0.356
CHA	0.177	516.400	91.566	43.701	47.865	97.80	97.524	2.465	-0.235	2.700	9.680	8479	8256	8483.840	8251.160	48.514	48.833	0.318
GSW	0.166	768.533	127.339	65.038	62.301	101.65	131.935	4.459	0.482	3.977	14.781	9421	8539	9428.391	8531.609	65.350	65.652	0.302
LAC	0.176	536.900	94.728	45.436	49.292	97.970	100.607	2.756	0.630	2.126	9.409	8569	8218	8573.705	8213.295	52.601	52.895	0.294
DAL	0.214	332.600	71.288	28.147	43.141	96.400	86.642	1.557	0.416	1.142	7.250	8388	8413	8391.625	8409.375	40.151	40.398	0.246
POR	0.117	734.817	86.338	62.185	24.153	98.310	49.469	2.030	0.911	1.120	4.129	8622	8554	8624.064	8551.936	43.256	43.392	0.137
SAS	0.236	225.183	53.215	19.056	34.158	95.720	68.117	2.990	1.020	1.970	6.264	8490	7618	8493.132	7614.868	67.133	67.264	0.132
MEM	0.069	456.833	31.728	38.660	-6.932	95.650	-13.814	1.138	0.243	0.895	-1.122	8126	8310	8125.439	8310.561	34.666	34.628	-0.038

Practical implications

While I hope that this investigation will help improve coaches' substitution decisions, I am keenly aware of the typical delay between these kinds of analyses and their eventual adoption in professional sports. For example, in 2006, David Romer provided strong evidence that NFL coaches, when faced with a short fourth down, should go for it (i.e. not punt) almost every time (Romer, 2006). A decade later in the 2015-2016 seasons, in situations in which punting is most costly¹⁴, coaches still opted to punt almost 60% of the time. Nonetheless, that number has steadily decreased since the publication of Romer's analysis (Stuart, 2017).

That said, if the NBA as a whole is slow to adopt this strategic adjustment, the door is open for one or two teams to gain a considerable competitive advantage. In the NFL, the 2017 Philadelphia Eagles analogously used fourth down attempts to gain an edge, and ultimately won the Super Bowl. They attempted 26 fourth down conversions, which ranked second in the league, and converted two very memorable fourth down attempts in the Super Bowl en route to defeating the heavily favored New England Patriots (Shpigel, 2018; Bonesteel, 2018). Challenging the status quo may be jarring, but the first mover has the most to gain.

Limitations and future directions

One aspect that my analysis does not address is the emotional impact of a player's disqualification. It is possible that if a player, particularly a star player, fouls out of a game, it might be deflating for the rest of the team. Answering this kind of question is always difficult, but it is especially tricky in the present investigation. The only way to measure the emotional impact of a player fouling out would be to compare to similar situations in which that player did

¹⁴ These are plays that occur in the first three quarters, between the opponent's and one's own 40-yard line, and when the score is within 10 points.

not foul out—that is, minutes at the end of a game when a star player is not in the game, but has not fouled out. This situation does not happen frequently enough to provide sufficient data. Furthermore, a positive emotional bump from a star player’s disqualification seems just as plausible. Anecdotally, teams often rally around each other when they are missing a star player. It is therefore unclear what effect, if any, the inclusion of the emotional effect of a foul-out might have on the analysis.

Another potential limitation is that these analyses use data from just the 2015-2016 season. This is primarily because several of the analyses rely on end-of-season statistics. For example, in both Section 2 and Section 3, I use players’ end-of-season VORP in parts of the analyses. That said, a single season provides ample data for my investigation, and there is no reason to think that that particular season is unique.

And finally, my analysis is primarily focused on aggregate effects, and so there is certainly room for individual differences. I cannot point to any single decision and determine whether or not it is the right one. It is possible that any one specific player might hurt his team when he is in foul trouble or might be substantially more valuable at the ends of games. Future work might delve more deeply into these individual differences and provide specific advice for certain situations. Nonetheless, the present analysis demonstrates that overall, coaches would be well served by allowing players to risk fouling out themselves, rather than artificially doing it for them.

References

Benartzi, S., & Thaler, R. H. (1995). Myopic loss aversion and the equity premium puzzle. *The quarterly journal of Economics*, 110(1), 73-92.

Bonesteel, M. (2018, February 5). 'Our coach has got some guts, huh?': Doug Pederson never backs down in Eagles' Super Bowl win. *The Washington Post*. Retrieved from www.washingtonpost.com.

Donaghy, T. (2010). *Personal foul: A first-person account of the scandal that rocked the NBA*. Clerisy Press.

Jacobs, J. (2017, September 18). Deep dive on regularized adjusted plus minus II: Basic application to 2017 NBA data with R. [Blog post]. Retrieved from <https://squared2020.com/2017/09/18/deep-dive-on-regularized-adjusted-plus-minus-ii-basic-application-to-2017-nba-data-with-r/>

James, B. (1983). *The Bill James Baseball Abstract 1983*. New York, NY: Ballantine Books.

Jennings, D. L., Amabile, T. M., & Ross, L. (1982). Informal covariation assessment: Data-based versus theory-based judgments. In D. Kahneman, P. Slovic, & A. Tversky (Eds.), *Judgment under uncertainty: Heuristics and biases* (pp. 211-230). Cambridge University Press.

Johns, D. M. (2011, March 7). The Great Carmelo Debate. [Blog post]. Retrieved from http://www.slate.com/articles/sports/sports_nut/2011/03/the_great_carmelo_debate.html

Kahneman, D., & Miller, D. T. (1986). Norm theory: Comparing reality to its alternatives. *Psychological review*, 93(2), 136.

Kahneman, D., & Tversky, A. (1979). Prospect Theory: An Analysis of Decision under Risk. *Econometrica*, 47(2), 263-292.

Kahneman, D., & Tversky, A. (1984). Choices, values, and frames. *American Psychologist*, 39(4), 341.

Larsen, A. (2013, January 31). The Offbeat: A Quick Note on Individual Defensive Ratings. [Blog post]. Retrieved from <https://www.slcduke.com/2013/1/31/3939352/the-offbeat-a-quick-note-on-individual-defensive-ratings>

March, J. G., & Shapira, Z. (1987). Managerial perspectives on risk and risk taking. *Management Science*, 33, 1404-1418.

Maymin, A., Maymin, P., & Shen, E. (2012). How Much Trouble Is Early Foul Trouble? Strategically Idling Resources in the NBA. *International Journal of Sport*

Finance, 7(4), 324-339.

Miller, D. T., & Taylor, B. R. (1995). Counterfactual thought, regret, and superstition: How to avoid kicking yourself. *What might have been: The social psychology of counterfactual thinking*, 305-331.

Morey, D. in Dewan, J., & Zminda, D. (1993). *STATS Basketball Scoreboard 1993-1994*. Harper Perennial.

Moskowitz, T., & Wertheim, L. J. (2011). *Scorecasting: The hidden influences behind how sports are played and games are won*. Three Rivers Press (CA).

Oliver, D. (2004). *Basketball on paper: rules and tools for performance analysis*. Potomac Books, Inc..

Parker, R. J. (2009a, October 25). Individual offensive efficiency ratings extracted from play-by-play data. [Blog post]. Retrieved from <http://www.basketballgeek.com/2009/10/25/individual-offensive-efficiency-ratings-extracted-from-play-by-play-data/>

Parker, R. J. (2009b, October 30). Individual defensive efficiency ratings extracted from play-by-play data. [Blog post]. Retrieved from <http://www.basketballgeek.com/2009/10/30/individual-defensive-efficiency-ratings-extracted-from-play-by-play-data/>

Risen, J. L., & Gilovich, T. (2008). Why people are reluctant to tempt fate. *Journal of Personality and Social Psychology*, 95(2), 293.

Romer, D. (2006). Do firms maximize? Evidence from professional football. *Journal of Political Economy*, 114(2), 340-365.

Rosenbaum, D. T. (2004, April 30). Measuring how NBA players help their teams win. [Blog Post]. Retrieved from <http://www.82games.com/comm30.htm>.

Shpigel, B. (2018, February 2). How the Eagles Followed the Numbers to the Super Bowl. *The New York Times*. Retrieved from www.nytimes.com.

Stern, H. S. (1994). A Brownian motion model for the progress of sports scores. *Journal of the American Statistical Association*, 89(427), 1128-1134.

Stuart, C. (2017, November 10). Smart Coaches Don't Punt. Retrieved April 9, 2018, from http://www.slate.com/articles/sports/sports_nut/2017/11/nfl_coaches_are_making_better_fourth_down_decisions.html.

Tango, T. M., Lichtman, M. G., & Dolphin, A. E. (2007). *The book: Playing the*

Percentages in Baseball. Potomac Books, Inc..

Thaler, R. (1985). Mental accounting and consumer choice. *Marketing science*, 4(3), 199-214.

Thompson, E. (2012, June 18). Durant's Foul Trouble Proves Costly to Thunder. *The New York Times*. Retrieved from www.nytimes.com.

Tversky, A., & Kahneman, D. (1973). Availability: A heuristic for judging frequency and probability. *Cognitive psychology*, 5(2), 207-232.

Tversky, A., & Kahneman, D. (1992). Advances in prospect theory: Cumulative representation of uncertainty. *Journal of Risk and uncertainty*, 5(4), 297-323.

Walco, D. K., & Risen, J. L. (2017). The empirical case for acquiescing to intuition. *Psychological science*, 28(12), 1807-1820.

Walker, J., Risen, J. L., Gilovich, T., & Thaler, R. (2018). Sudden-death aversion: Avoiding superior options because they feel riskier. *Journal of Personality and Social Psychology*.

Weinstein, J. (2010a, April 27). Foul trouble [Blog post]. Retrieved from <https://theoryclass.wordpress.com/2010/04/27/foul-trouble/>.

Weinstein, J. (2010b, May 17). Foul follow-up [Blog post]. Retrieved from <https://theoryclass.wordpress.com/2010/05/17/foul%C2%A0follow-up/>.

Appendix 1: List of star players

Star players are defined as any player who made an all-star team between the 2013-2014 season and the 2015-2016 season:

LaMarcus Aldridge	Tim Duncan	Dwight Howard	Joakim Noah
Carmelo Anthony	Kevin Durant	Kyrie Irving	Dirk Nowitzki
Chris Bosh	Marc Gasol	LeBron James	Tony Parker
Kobe Bryant	Pau Gasol	Joe Johnson	Chris Paul
Jimmy Butler	Paul George	Kyle Korver	Jeff Teague
DeMarcus Cousins	Draymond Green	Kawhi Leonard	Isaiah Thomas
Stephen Curry	Blake Griffin	Damian Lillard	Klay Thompson
Anthony Davis	James Harden	Kevin Love	Dwyane Wade
DeMar DeRozan	Roy Hibbert	Kyle Lowry	John Wall
Andre Drummond	Al Horford	Paul Millsap	Russell Westbrook

Appendix 2: Offensive rating details

- If there is an offensive rebound in a possession that results in points scored, the offensive rebounder receives a portion of the credit for the points and possession proportional to:

$$\frac{(1 - TeamOR\%)(TeamPlay\%)}{(1 - TeamOR\%)(TeamPlay\%) + (TeamOR\%)(1 - TeamPlay\%)}$$

where TeamOR% is the team's overall offensive rebound rate (offensive rebounds/(offensive rebounds + defensive rebounds)), and TeamPlay% is the percent of offensive possessions that result in at least one point being scored (scoring possessions/total possessions).

- If a shot is made without an assist, the shooter receives full credit for the points and the possession.
- If there was an assist, the assister receives a portion of the credit based on where on the court the shot was taken from. The logic is essentially that the closer to the basket a made shot is, the more credit the assister deserves and the less credit the shooter deserves. More specifically, the assister receives credit proportional to half of the effective field goal percentage (eFG%) of the shot. I divide shots into five common categories: restricted area (< 4 feet from the basket), paint (inside the painted area), other two-point shots (any non-three that is outside the paint), corner three (a three point shot taken from the corner, which is closer than other threes), non-corner three (a three-point shot not taken from the corner). The eFG% is calculated as the field goal percentage for a shot from each of these zones, multiplied by the points a shot is worth from that zone (i.e. either two or three).
- On a free throw, the shooter receives full credit for the points and possession. If the free throw is part of an and-1 and the basket was assisted, the assister receives a portion of the

credit for the possession and the points on the made shot, but not for the point on the free throw.

- On a missed shot that is rebounded by the defense, the shooter receives full credit for the possession.
- On a turnover that is assigned to a specific player, he receives full credit for the possession. If it is a team turnover, all players on offense receive 20% of the possession.

Appendix 3: Defensive rating details

- On a made shot that is not part of an “and-1”, it is not clear which player on defense was guarding the shooter. Therefore, all players on the court for the defense are credited with 20% of the points scored and 20% of the possession. If the made shot is part of an and-1, the foul is credited with the points.
- On a missed shot that was blocked and was rebounded by the defense, the credit for the possession is split between the blocker and the defensive rebounder. The credit is weighted by the relative difficulty between forcing the missed shot and obtaining the defensive rebound. If the shot was not blocked, each player on the defense receives 20% of that portion of the credit. If the rebound was not credited to a single player (i.e. a team rebound), each player on the defense receive 20% of the credit for the rebound.
- On a made free throw, the player who committed the foul is credited with the point. If it's the last free throw and it goes in, the foul is also credited with the possession. If the foul was not called on a specific player (e.g. technical foul on the bench), all players on the court for the defense receive 20% of the credit.
- If the last free throw is missed, the possession is split between the defensive rebounder and the foul.
- On a steal or offensive foul, the player who stole the ball or drew the offensive foul is credited with the possession. On any other turnover, each player on the defense receives 20% of the credit for the possession.

Appendix 4: Section 2 robustness checks

Table 10: Robustness checks for regression model in Section 2

	(1)	(2)	(3)	(4) ^a	(5) ^a	(6)	(7)	(8) ^b	(9)	(10) ^c	(11) ^d
home.ft	1.058 (3.331)	5.702* (3.399)	6.194* (3.677)					3.637 (5.136)	5.543 (4.127)	5.437 (3.747)	6.207* (3.665)
away.ft	2.957 (2.957)	-4.116 (2.967)	-8.897*** (3.307)					-6.198 (4.598)	-6.737* (3.504)	-7.220** (3.347)	-8.913** (3.994)
home.1ft				5.084 (3.434)	5.358 (3.717)	4.952 (3.424)	5.677 (3.703)				
home.2plusft						25.105 (15.379)	21.617 (23.654)				
away.1ft				-4.794 (3.063)	-8.482** (3.390)	-4.998 (3.042)	-8.931*** (3.383)				
away.2plusft						10.670 (9.827)	-10.400 (12.226)				
VORP.dif	0.575*** (0.094)	0.681*** (0.088)	0.577*** (0.094)	0.679*** (0.088)	0.574*** (0.094)	0.680*** (0.088)	0.879*** (0.140)	0.672*** (0.140)	1.873*** (0.088)	1.873*** (0.168)	0.543*** (0.111)
VORP.dif.ft	-1.000 (1.271)	-1.382 (1.660)	-0.987 (1.339)	-0.975 (1.793)	-1.172 (1.266)	-1.542 (1.678)	-3.167 (2.436)	-0.935 (1.670)	-1.139 (1.661)	-1.069 (1.772)	
score.dif	1.929*** (0.054)	2.064*** (0.053)	1.928*** (0.054)	2.063*** (0.053)	1.931*** (0.054)	2.066*** (0.053)	2.039*** (0.080)	2.067*** (0.054)	2.229*** (0.057)	2.026*** (0.064)	
star.home.ft									2.967 (9.849)		
star.away.ft									-13.055 (8.712)		
fixed effects											
category	Minute	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
possessions	Minimum	0	0	4	0	4	0	4	4	4	4
DF	37,004	36,994	23,686	36,818	23,621	36,992	23,684	12,876	23,684	23,628	23,686

Standard errors are reported in parentheses. * , ** , *** indicates significance at the 90%, 95%, and 99% level, respectively.

a: This excludes all observations with 2 or more players in foul trouble for either team.

b: This includes only stints that have an even number of possessions (i.e. same number of stints for the home and away team).

c: This includes fixed effects for home team and away team

d: This is a quantile regression using median values.

Table 11: Additional robustness checks for regression model in Section 2 (by game segment)

	(1)	(2)
home.ft	0.650 (4.576)	12.082** (5.830)
away.ft	-13.750** (4.241)	-3.957 (5.436)
VORP.dif	0.386*** (0.098)	1.199*** (0.237)
VORP.dif.ft	0.880 (2.409)	-5.233** (2.192)
score.dif	2.724*** (0.071)	1.099*** (0.083)
Minute category fixed effects	Yes	Yes
Minimum possessions	4	4
Minute categories	1.1-3.2	4.1-4.2
DF	17,482	24,103

Standard errors are reported in parentheses. *, **, *** indicates significance at the 90%, 95%, and 99% level, respectively.

Appendix 5: Effect of players' skill difference on outcome, by game segment

Table 12: Regression results of score differential on various measures of player quality, by game segment

	Overall	1.1	1.2	2.1	2.2	3.1	3.2	4.1	4.2	4.2(close)
VORP.dif ^a	1.594*** (0.095)	1.827*** (0.370)	2.037*** (0.237)	2.759*** (0.434)	0.883*** (0.223)	1.910*** (0.360)	1.258*** (0.229)	2.267*** (0.422)	1.557*** (0.241)	0.446 (0.422)
Max.VORP.dif ^a	2.587*** (0.118)	3.108*** (0.839)	3.000*** (0.448)	4.207*** (0.815)	1.449** (0.458)	2.927*** (0.775)	2.192*** (0.434)	4.164*** (0.763)	2.559*** (0.494)	1.138 (0.842)
WS.dif ^b	0.740*** (0.045)	0.857*** (0.185)	1.005*** (0.114)	1.225*** (0.206)	0.411*** (0.109)	0.875*** (0.177)	0.574*** (0.109)	1.005*** (0.109)	0.698*** (0.115)	0.190 (0.202)
PER.dif ^c	0.654*** (0.045)	0.852*** (0.201)	0.923*** (0.115)	0.743*** (0.177)	0.460*** (0.107)	0.858*** (0.191)	0.560*** (0.109)	0.756*** (0.172)	0.538*** (0.109)	0.177 (0.210)
DF	36,220	880	6,546	3,995	6,367	1,499	6,463	3,902	6,183	2,364

Standard errors are reported in parentheses. *, **, *** indicates significance at the 90%, 95%, and 99% level, respectively.

a: It is possible that the quality of the best player on the floor for each team is more meaningful than the total quality of all players. Max.VORP.dif measures the difference in VORP between the best player for the home team and the best player for the away team.

b: WS.dif is difference in Win Shares.

c: PER.dif is difference in Player Efficiency Rating.

Appendix 6: Foul out simulation details

In order to estimate the expected minutes missed due to foul outs if coaches did not bench their players in foul trouble, I ran 10,000 simulations of a single starter's game. In each simulation, I used the following process:

1. I randomly sampled total minutes played for a single game from the distribution of all games played by starters in the 2015-2016 season. I excluded games in which starters played fewer than 10 minutes, as those missed minutes were likely due to injuries.
2. I assumed that players' playing time would be approximately equally divided between the four quarters. Specifically, for quarters one through three, I rounded the expected minutes played divided by 4, to the nearest whole minute. Then I assumed that the player played the remaining expected minutes in the fourth quarter. I assumed that the player played his minutes at the beginning of the first and third quarters, and at the end of the second and fourth quarters. This follows a typical pattern for how starters' minutes are distributed.
3. I simulated each minute of the player's game, such that his average foul rate when not in foul was 0.089, and his average foul rate when in foul trouble was 0.055.
4. If a player fouled out in the simulation (i.e. recorded 6 fouls), his final playing time was calculated as the total number of minutes he had played up to that point.
5. I recorded the player's expected minutes played (from Step 1), the number of minutes he was in foul trouble, and the number of minutes he ended up playing (the same as expected minutes unless he fouled out).