Foul Play:
Examining NBA Player Behavior in Response to Foul Trouble

Gabriel Zanuttini-Frank
Advisor: Jay Emerson

1 Research Question

- The rule of fouling out in the NBA, that a player who commits six fouls over the course of a game is disqualified for the remainder of the game, gives rise to very interesting strategic questions:
- Coaches must decide when to substitute a player given his foul situation, and players can adjust how aggressively they play to optimize the balance between committing more fouls and playing effectively.
- This project performs a statistical analysis of players' defensive tendencies to determine whether foul trouble affects the way players defend.
- **Hypothesis:** Players' defensive aggressiveness will respond negatively to foul trouble - as players get into foul trouble, they will play less aggressively on defense

2 Defining Player Aggressiveness

- I use player aggressiveness as a way of measuring players' defensive tendencies
- I define defensive aggressiveness as the distance from a defensive player to the closest offensive player
- SportVU data from the 2015-16 NBA season was used to conduct this analysis, in conjunction with basic play-by-play data

player_id	lastname	firstname	position	team_id	x_loc	y_loc	game_clock	shot_clock	quarter	event.id
200768	Lowry	Kyle	G	1610612761	10.77643	38.39262	713.26	13.15	1	2
200768	Lowry	Kyle	G	1610612761	10.77643	38.39262	713.26	13.15	1	3
200768	Lowry	Kyle	G	1610612761	10.77643	38.39262	713.26	13.15	1	4

Table 1: Sample of SportVU player tracking dataset

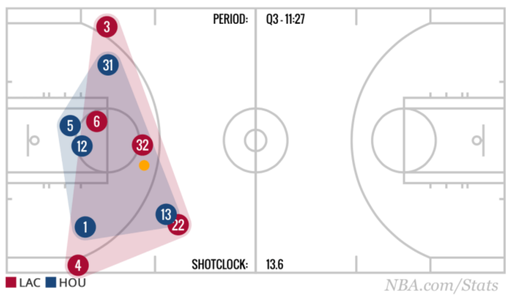

Figure 1: Example of visual created from player tracking movement data

3 Defining Foul Trouble

- Foul trouble is loosely defined as a situation in which a player is in danger of exceeding the number of permitted fouls and this being disqualified from the game
- Potential quantitative measures of foul trouble:
  - Number of fouls
  - Squared number of fouls
  - Probability of fouling out
- I model the likelihood of a player fouling out given his number of fouls and the time remaining in the game using a Markov chain. Transition probabilities were estimated using the time between each number of fouls from 150 NBA games

Table 2: Transition matrix for fouling at any second

	0	1	2	3	4	5	6
0	0.99909	0.00091	0.00000	0.00000	0.00000	0.00000	0.00000
1	0.00000	0.99926	0.00074	0.00000	0.00000	0.00000	0.00000
2	0.00000	0.00000	0.99923	0.00077	0.00000	0.00000	0.00000
3	0.00000	0.00000	0.00000	0.99916	0.00084	0.00000	0.00000
4	0.00000	0.00000	0.00000	0.00000	0.99924	0.00076	0.00000
5	0.00000	0.00000	0.00000	0.00000	0.00000	0.99934	0.00066
6	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	1.00000

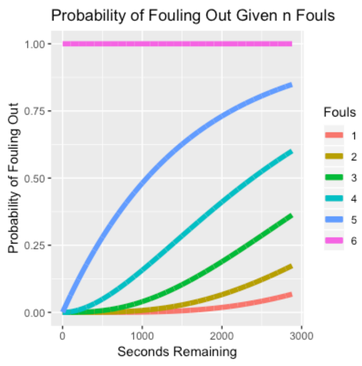

Figure 2: Illustration of the probability of fouling out before the end of the game, given n fouls and the time remaining

4 Data Cleaning

- One of the most challenging problems in this project was to merge the player tracking and play by play datasets, as there were not variables on which to merge reliably
  - Merged using significant changes in the shot clock to denote ends of plays
- Many issues were found in the data; some missing data or slight inconsistencies, and some significant inaccuracies in measurement
- Filtered the dataset to remove duplicate rows of (player ID, game clock), and kept only observations of the five players currently on defense.
- Final dataset used for analysis was ~10% the size of the initial

5 Analysis

- Validating distance to closest opponent as a measure of aggressiveness:

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	6.7835	0.0089	758.55	0.0000
scale(total_clock)	-0.0178	0.0096	-1.86	0.0626
scale(abs_scorediff)	0.0552	0.0095	5.79	0.0000
scale(shot_clock)	0.4286	0.0090	47.70	0.0000

Table 3: Predicting defensive distance on jump shots using time remaining, absolute value of score differential, and shot clock

- Score differential and shot clock are thought of as indicators of times of more or less aggressive play, and the signs of their coefficients point in the intuitive direction: later in the shot clock and in closer games, the defense will play more tightly
- Total clock was not significant at a significance level of .05. Further, we would expect its coefficient to be positive. I will proceed without this variable.
- Playing tighter defense equates to being more aggressive because it increases likelihood of fouling:

Number of Fouls	Did he foul on this play?	Mean Distance
0	FALSE	5.48
0	TRUE	4.36
1	FALSE	5.45
1	TRUE	4.37
2	FALSE	5.43
2	TRUE	4.26
3	FALSE	5.41
3	TRUE	4.36
4	FALSE	5.58
4	TRUE	4.14
5	FALSE	5.70
5	TRUE	3.66

	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	-2.4962	0.0188	-132.95	0.0000
closest_dist	-0.1259	0.0037	-34.36	0.0000

Table 4: predicting whether the player guarding the ball committed a foul on a given play using his defensive distance

- Which of the three foul trouble measures best predicts defensive aggressiveness?

Table 5: Regression using number of fouls

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	1.4506	0.0020	732.42	0.0000
scale(abs_scorediff)	0.0083	0.0020	4.13	0.0000
scale(shot_clock)	0.1200	0.0020	60.51	0.0000
scale(num_fouls)	-0.0030	0.0020	-1.50	0.1330

Table 6: Regression using squared number of fouls

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	1.4506	0.0020	732.42	0.0000
scale(abs_scorediff)	0.0079	0.0020	3.96	0.0001
scale(shot_clock)	0.1201	0.0020	60.58	0.0000
scale(sq_fouls)	-0.0004	0.0020	-0.18	0.8550

Table 7: Regression using the modeled probability of fouling out

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	1.4506	0.0020	732.42	0.0000
scale(abs_scorediff)	0.0077	0.0020	3.89	0.0001
scale(shot_clock)	0.1201	0.0020	60.64	0.0000
scale(foulout_prob)	-0.0012	0.0020	-0.61	0.5399

- None of the three show significant results
  - Restricting data by position, type of shot, or quarter did not improve results

6 Discussion

- Despite many attempts to subset the data and various definitions of foul trouble, no significant evidence was found that a player's foul trouble status affects his distance from his matchup on defense
- Conclusions: either the measure of aggressiveness used does not capture variations in defensive play in response to foul trouble or players are simply not adjusting their play to foul trouble.
- This analysis could be improved by using alternative measures of aggressiveness that are not available in this data, such as hand placement and jump frequency near the rim. We could also use other measures to assess whether players adjust their defensive behavior, such as quantifying their defensive effectiveness.
- Used 144 games of data, but distances are extremely noisy and subject to a huge number of variables
  - R-squared values were extremely low, never above .05