Quantifying Foul "Tilt" of NBA Players

Katherine Evans, PhD

Verily Life Sciences

CMU Sports Analytics Conference 2018

Disclaimer

Opinions expressed are solely my own and do not express the views or opinions of my employer

Introduction

What Do I Mean By “Tilt?”

- Tilt is a term that originated in poker and refers to a general state of being where a person is confused or frustrated and then adopts a suboptimal strategy
  - Typically over-aggression.
- Over-aggression then often leads to more suboptimal play which leads to more frustration.

- How could we measure tilt?

Boogie Sure Does ~~Get Upset~~ Foul A Lot

Season	Games	Fouls
10-11	81	332
11-12	64	257
12-13	75	269
13-14	71	270
14-15	59	241
15-16	65	236
16-17	72	278
17-18	48	183

Do Fouls Beget Fouls?

- Once a player commits a foul, is he more likely to commit another one?
- Moreover, is a player more likely to commit a 5^{th} foul once he has 4 than he is to commit a 2^{nd} foul one he has 1?

Modeling

- We can consider a survival model, and look at the “failure time” for each foul
  - ie. the time it takes a player to commit his 1st found, 2nd foul, etc.
- If, for example, the time between the 2nd and 3rd foul is significantly longer than than the time between the 4th and 5th foul, we might have evidence of some sort of “tilt.”
  - Player time, not game clock time
- We can use a stratified Cox model to analyze the data.
- Also, going to ignore technical fouls for now
  - Which are probably a bigger indicator of “tilt”
  - But way less frequent

Data

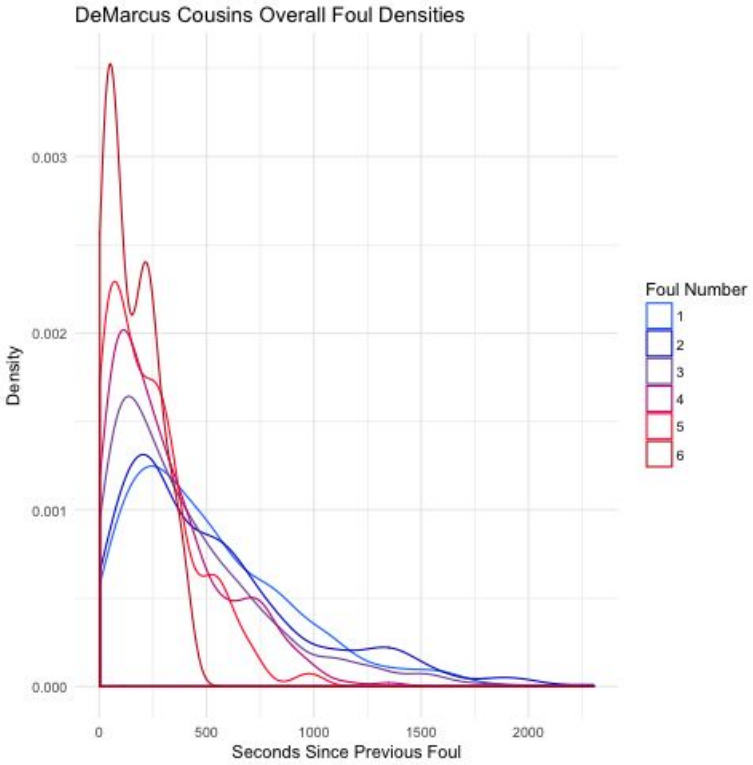
Data

- Play-by-play data and box-score data from the NBA for the 2011-2012 through 2017-2018 seasons.
  - This data is publicly available from nba.com
  - Cleaned up data available at [eightthirtyfour.com]
- Using the box-score data and substitutions in the play by play for each game, we can determine the amount of time any given player has actively played in the current game at each event in the play by play data.
- We also generated censoring times using the maximum playing time for a given player
  - This is because games end, censoring observations

Basic Analysis

Censoring

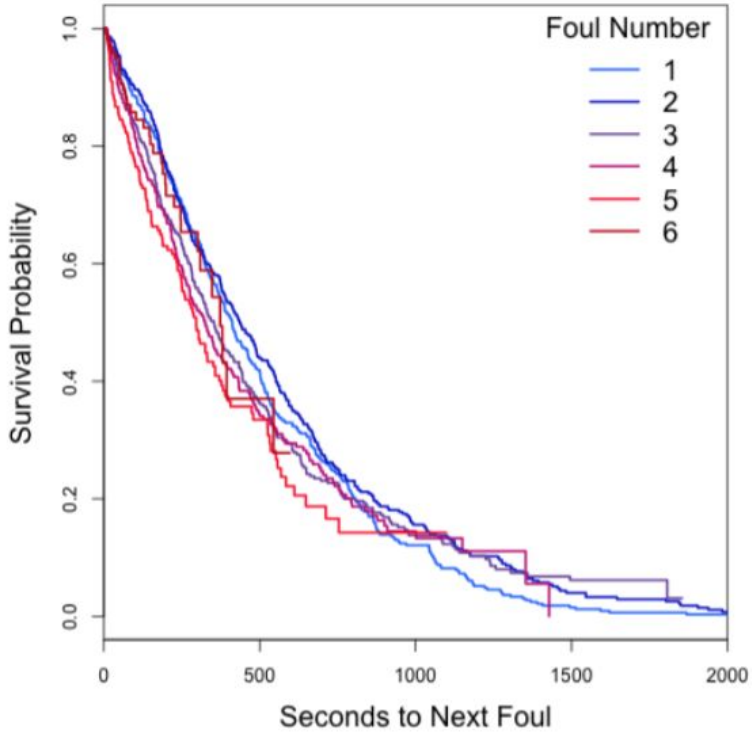
- If games were infinitely long, and players continued to play, we would observe every player until he fouled out.
- Many fouls are censored due the end of follow up time.
- Censoring biases the numbers we see
    - The graph looks like it supports our hypothesis, but the underlying data is biased because of censoring



Rate densities when we ignore censoring

Kaplan-Meier Curves

- A survival curve, in general, is used to map the length of time that elapses before an event occurs.
- Here, they give the probability that a player has “survived” to a certain time without committing a particular number foul.
- The graph is now less biased
    - But still problematic



KM Curves for DeMarcus Cousins

Basic Summary

- Basic summary statistics give some insight
  - Time to foul seems to shrink
- But again, these numbers do not account for censoring
  - Eg. Boogie had 72 games where he committed only 1 or 2 fouls
- **We need to look only at games with a large number of fouls**

DeMarcus Cousins			
Total Fouls	Mean Time	Games	# Games with at least this many fouls
1	502.36	16	440
2	516.21	56	424
3	397.45	95	368
4	310.46	116	273
5	236.45	115	157
6	151.07	42	42

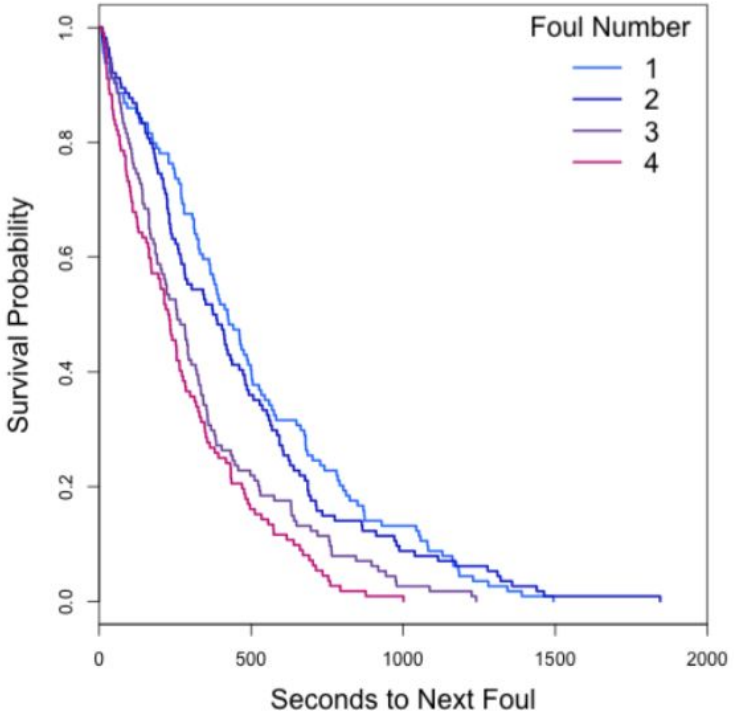
Subset the Data

- Instead, we will examine games for each player where they committed a minimum of five fouls
  - And limit our analysis to the first four fouls.
  - Do this to ensure we have comparable games
  - Because of the censoring...
- This foul restriction gives us a larger sample size, though restricts us from gaining understanding about how players accrue their 5th and 6th fouls.

Better Analysis

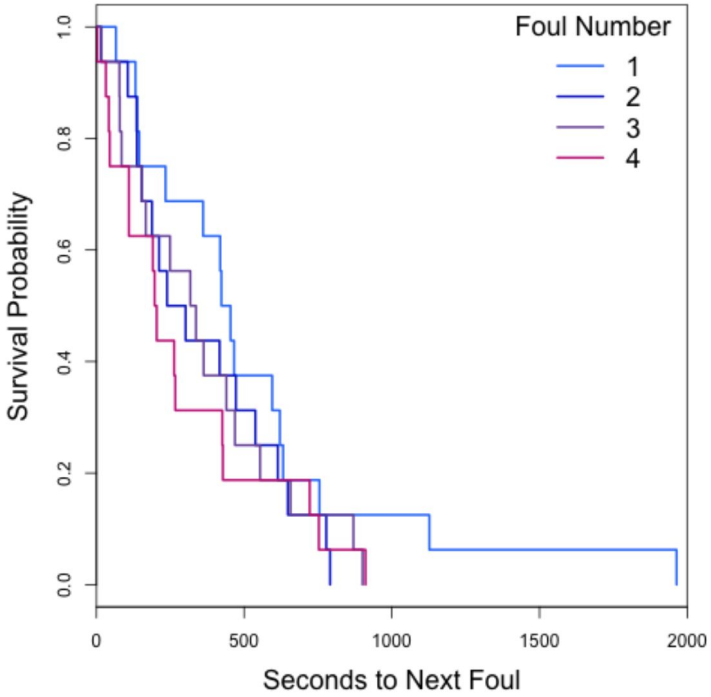
Kaplan-Meier Curves Part 2

- Now we can see pretty distinct ordering in how fouls are accrued
- If there were no “tilt,” then we’d expect to see fouls occur a lot more randomly
- Evidence towards our hypothesis
    - Though certainly not proof

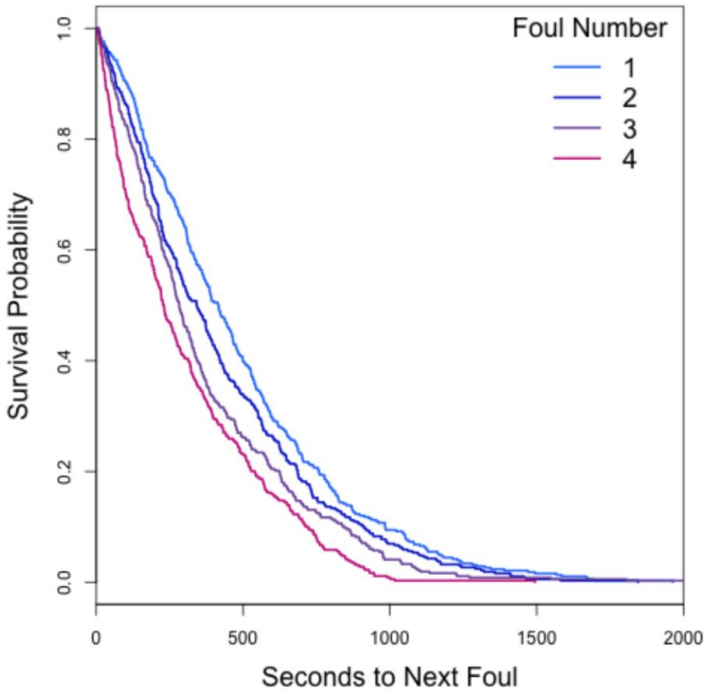


KM Curves for DeMarcus Cousins

KM Curves for More Players



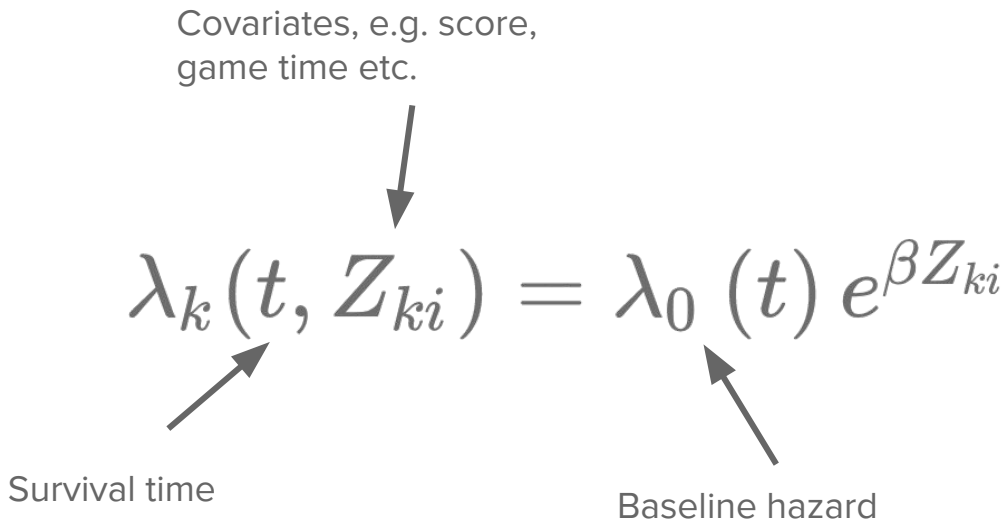
KM Curves for Al Horford



KM Curves for Cousins, Horford, Howard, Bogut, Chandler, M. Gasol, B. Lopez, & R. Lopez pooled

Time for Some Math

- Conditional risk set model since we have repeated events/fouls
  - But really a stratified Cox model since we distinguish between fouls



Cox Model Results

- Let's also control from some other variables, score and game time
- We see that fouls 2, 3, and 4 are committed sooner than the prior foul.
- Hazard ratio = $\exp(\text{coeff}_k - \text{coeff}_{k-1})$
  - Eg. when Cousins has 3 fouls he is 374% times as likely to commit a foul at any given time than when he only has 2 fouls.

	DeMarcus Cousins
Score Diff	0.01 (0.01)
Game Time*Score Diff	−0.0003 (0.0004)
Game Time	−0.11*** (0.01)
Foul 2	0.82*** (0.12)
Foul 3	2.14*** (0.16)
Foul 4	3.21*** (0.20)
Observations	628

Estimate (Standard Error)

Further Thoughts

- Oh god, its so much more complicated
- All this subsetting allows for some comparable fouls, but then we are ignoring all the games where he didn't foul very much
- What about hacking?
- What about the other side of the court?
    - Players may get mad if they are fouled
- Censoring makes these sorts of problems really hard

Thank You!

- Udam Saini - my amazing collaborator
  - All our data is available at eightthirtyfour.com
- Sam - for tricking me into giving this talk
- My job - for having an office nearby and for letting me do sports as a side gig
- Mike Lopez - even though he's not here
- Jackie Bradley Jr - for nine RBIs with two outs

And, of course,

- DeMarcus Cousins - a constant source of inspiration and entertainment

Thanks!

Contact me:

causalkathy@gmail.com

@CausalKathy on Twitter

www.CausalKathy.com