A Multilevel Bayesian Approach for Inferring the Effect of Fouls on Offensive Patterns in NBA Basketball Games

AUTHOR
Michael Light

Timeline

Editable here.

date	task
Thu, Apr 10, 2025	First Generative Model Validated w/SBC
Fri, Apr 11, 2025	First Computational Model Fit to Real Data
Mon, Apr 14, 2025	Second Generative Model Validated w/SBC
Wed, Apr 16, 2025	Second Computational Model Fit to Real Data
Fri, Apr 18, 2025	Final Generative Model Validated w/SBC
Mon, Apr 21, 2025	Final Computational Model Fit to Real Data
Thu, Apr 24, 2025	Initial Draft Ready for Review
Mon, Apr 28, 2025	Final Draft Submitted

Project Motivation

Background on the Role of Physicality in NBA Games

Basketball is fundamentally a simple game: put a ball through a hoop. It's also a numbers game: put a ball through a hoop more times than your opponent. As a result, basketball analysis has historically focused on offensive aspects - who scored the most points, how they scored them, who assisted them, and so on. This offensive focus is natural given that the final score directly reflects which team scored more points.

The introduction of player tracking technology in the early 2010s briefly revolutionized defensive analysis by providing detailed data on defender positioning and movement. However, since 2016, this valuable player tracking data has been made private, creating a significant challenge for researchers outside NBA organizations.

This data limitation has created a critical gap in basketball analytics. Without access to detailed player tracking data, researchers are left with the same fundamental problem they faced before the tracking era: how can we effectively study and evaluate individual defensive ability when publicly available statistics are either: - Primarily focused on offense - Only loosely related to defense (e.g., blocks and steals)

Solving this problem would democratize defensive evaluation by creating metrics that can be derived from publicly available, play-by-play data. This would allow analysts outside NBA organizations to contribute meaningful insights to the game and help properly communicate the value of defensive players who may not show up in traditional box scores.

At a broader level, developing robust defensive metrics from public data could help bridge the growing divide between basketball analytics and fan understanding. Currently, there exists a significant gap between analysts who rely on advanced metrics and fans who primarily use traditional counting stats like points, rebounds, and assists. Developing novel, simple, and intuitive ways of understanding the impact of difficult-to-measure latent player characteristics like defensive skill represents a crucial challenge in modern basketball analysis.

Debate Surrounding Defense, Existing State of Research

Project Scope and Significance

Relevance and Novelty

Broader Context

Project Research Questions

Primary Research Question

Is there a detectable effect of defensive players’ foul accumulation on the shot-taking patterns of an opposing offense?

Secondary Research Questions

Is there a detectable effect of defensive players’ foul accumulation on the degree of shot pressure experienced by an opposing offense?

Literature Review

Bayesian models have been used in a number of previous studies to model the relationship between offensive and defensive performance. Williams et al. (2024) used a Bayesian hierarchical modeling framework to model offensive performance of teams and players based on their shooting propensities and abilities. They used discretized shot loactions for the top 100 shot-takers from the 2008/2009–2020/2021 seasons to create an expected points above average (EPAA) metric. Our project differs in our use of hierarchical modeling to infer defensive impact, rahter than contributing to the existing research on offense.

Franks et al. (2015) was one of the pioneering papers in using Bayesian hierarchical models to study player tracking data by modeling the relationship between a defensive assignment and the spacial-temporal shotmaking tendencies of various offensive players, which is to say how well they shoot the ball given the location, time, and defender of a shot attempt. Our project differs in that we use a Bayesian hierarchical model to infer defensive impact without using raw player tracking data, rather than modeling the relationship between a known defensive assignment and shotmaking tendencies.

Finally, Chu and Swartz (2020) use Bayesian methods to model the time-to-event of a foul as a gamma distributed function of a player, their accumulated number of fouls, and their position. Their goal was to develop a decision-making framework for coaches to use in determining when to sit players at risk of fouling out (usually when a player reaches 6 fouls) of a game. Although similar in spirit, our project doesn't model the time-to-event of a foul nor attempt to help coaches make decisions about sitting their players, but instead focuses on the effect of fouls while players are on the court.

Data Sources

Primary Data Sources

NBA Play-by-Play

- **Years:** 2023-2025
- **Size:** tk
- **Link:**

NBA Closest Defender Shooting Dashboard (player tracking)

- **Years:** 2023-2025
- **Size:** tk
- **Link:**

Causal Models

Simple Causal Model of Foul Effect on Shot-taking

More Complex Causal Model of Foul Effect on Shot-taking

More Complex Causal Model of Foul Effect on Shot-taking with Multiple Observation Processes

Generative Models

Simple Generative Model of Foul Effect on Shot-taking

More Complex Generative Model of Foul Effect on Shot-taking

More Complex Generative Model of Foul Effect on Shot-taking with Multiple Observation Processes

Statistical Models

Simple Statistical Model of Foul Effect on Shot-taking

More Complex Statistical Model of Foul Effect on Shot-taking

More Complex Generative Model of Foul Effect on Shot-taking with Multiple Observation Processes

Prior Selection and Predictive Checks

Two Prior Models for each Model of Foul Effect on Shot-taking

Less Constrained

More Constrained

Model Validation

Simulation Based Calibration

Observed Sample Selection

Observed Sample of Shots (Play-by-Play)

Observed Sample of Shots (Shot Chart Detail)

Observed Sample of Shots (player tracking)

Model Fitting and Diagnostics

Fitting Models to Observed Samples

HMC Diagnostics

Posterior Analysis

Posterior Retrodictive Checks

Posterior Residual Checks

Counter-factual Analysis

Conclusion

Appendices

Appendix A.

References

Chu, D., & Swartz, T. (2020). Foul accumulation in the NBA. Journal of Quantitative Analysis in Sports, 16(4), 301-309. https://doi.org/10.1515/jqas-2019-0119

Alexander Franks. Andrew Miller. Luke Bornn. Kirk Goldsberry. Characterizing the spatial structure of defensive skill in professional basketball. (2015) Ann. Appl. Stat. 9 (1) 94 - 121. https://doi.org/10.1214/14-AOAS799

Williams, B., Schliep, E. M., Fosdick, B., & Elmore, R. (2024). Expected points above average: A novel NBA player metric based on Bayesian hierarchical modeling. arXiv preprint arXiv:2405.10453. https://arxiv.org/abs/2405.10453