



THE UNIVERSITY OF CHICAGO

Department of Statistics

MASTER'S THESIS PRESENTATION

DANIEL CUNNINGHAM

Department of Statistics
The University of Chicago

Improved Major League Baseball Batting Projections Using a Random Forest Model

TUESDAY, October 30, 2018, at 1:45 PM
Jones 304, 5747 S. Ellis Avenue

ABSTRACT

Batting events in baseball are heavily influenced by luck and variance, making it difficult to predict future performance from past performance. With the advent of Statcast, it is possible to observe raw batted ball measurements that are less affected by luck and more directly related to a player's skill. In this talk, I present a random forest model used to predict the expected outcomes of batted balls based on Statcast measurements and other context information. The model can be used to identify players who have been lucky or unlucky over a period of time.

Furthermore, the predicted outcomes from the random forest model are more predictive of future performance than players' past performance. I present a simple framework for using the predictions from the random forest model to produce complete batting projections. The projections are shown to outperform the popular Marcel the Monkey Forecasting system over the 2017 and 2018 seasons. This exciting result shows that Statcast measurements can provide considerable value to player evaluation in Major League Baseball.