**Incorporating High School Recruitment Ratings and Statistics**

**in Predictive Models for Collegiate Basketball Success**

Ted Henson and Mario Giacomazzo

Accusations of many NCAA basketball coaches paying high profile recruits hundreds of thousands of dollars catalyzed this analysis. If one player can cause decorated coaches to potentially resort to unethical methods, then programs should explore every possible avenue of predicting college performance, especially programs with smaller budgets and less recognition. Many NCAA teams take many traditional approaches, i.e film studies, scouting, interviews, to make scholarship decisions. These methods can be time-consuming, expensive, and lead to subjective decision-making.

ESPN’s recruiting database assigns ratings to high school basketball players based on scouting reports. Since many universities rely on expert opinion from scouts, it is reasonable to conclude that recruitment ratings can be used to predict collegiate success. Two websites, prepcircuit.com and AAUStats.com, report box score statistics for high school basketball players. This data is publicly available online but unstructured requiring the use of web-scraping techniques.

Win shares is an advanced metric respected for its all-encompassing nature of a player’s contribution to team success. Based on the current “one-and-done” climate, universities are looking for players to contribute immediately and NBA teams are drafting many players without a full season of collegiate data; therefore, freshman season win shares was the response variable. The recruitment rankings and high school statistics provide inputs to linear and nonlinear models compared on predictive accuracy of the response variable. Given the novelty of the data we are analyzing, we use this as an opportunity to compare the impact that these different sources of data have on our predictive models. The best model had out-of-sample RMSE of 1.26 win shares. In this model, the incorporation of high school statistics improved out-of-sample prediction and the ESPN rating system that has been respected for 13 years. This finding will encourage more analysis and collection of high school data for collegiate and professional purposes.