`

Analyzing Load-Management Strategies: Age-Conditioned Average Treatment Effects and Bootstrapping for Professional Basketball

Raymond Lee, Sahil Singh, Dohyun Lee

Yale University

Introduction

In the dynamic world of professional sports, optimizing athlete performance through strategic load management has become a focal point of interest, particularly within the National Basketball (NBA) and Women's National Basketball Association (WNBA). The rigorous competition and physical demands in these leagues underscore the paramount importance of effective load management, striking a delicate balance between achieving peak performance and minimizing injury risks.

Our study employs sophisticated methodologies to explore age-related nuances in load management strategies. For WNBA athletes, we analyze age-conditioned average treatment effects (ACTE) curves, unraveling the interplay between age and the impact of load management on player performance. In addition, we also add a bootstrapping for estimation of errors in our ACTE curves of both NBA and WNBA athletes. Our investigations not only contribute to sports science discourse but also offer actionable insights for coaches and practitioners, aiming to refine age-specific load-management strategies in professional basketball.

Part 1: ACTE for WNBA Data

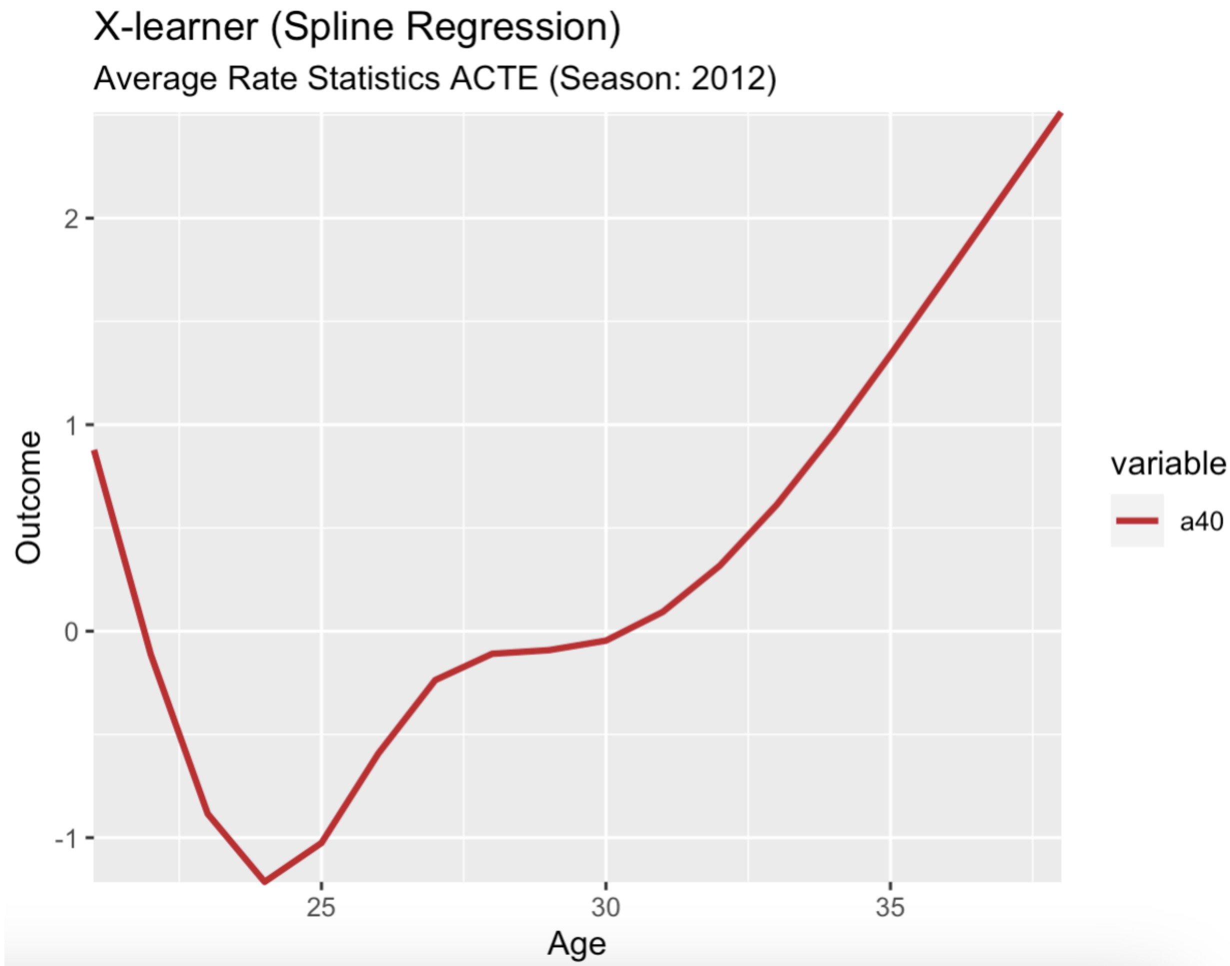
Data Preprocessing

To apply the ACTE functions on WNBA athletes, we used data regarding games from 2003 to 2022, player box scores, and player information available on the *wehoop* R package. We read in all the player box datasets and combined all the RDS files to get a comprehensive player box dataset. We followed the same process for the games data. The player IDs for each athlete in the player box data and games data don't match, so we had to find an alternative method to uniquely match the games data and player box data for each athlete in order to conduct a merge. To do this, we decided to create a new variable called "player_name_key" which would serve as the key ID. To clean the player names, we first implemented a function called *remove_middle_name.R*, which takes a full name string from the players dataset as the input parameter and returns the name with the middle name removed, retaining only the first and last names. We then implemented another function called *clean_player_name.R* which takes a string representing a player's name and performs cleaning operations using regular expressions to remove any non-alphabetic characters and convert the name to lowercase. The function also addresses any double spaces that might occur as a result of character removal, ensuring a standardized output. After merging the player box dataset and games dataset by the player name key and game ID, we found that there were still names that weren't affected by the merge, evident by the NA values that would be in player ID and player name. This was due to names that followed a unique, non-standard format such as Asian names that would have the order of the last name switched, names with typos in one dataset but correctly spelled in the other, names with foreign characters, or names with an athlete's legal first name in one dataset and their nickname in another. We then only kept rows with complete cases and converted the full team names to their three letter abbreviation. All of this is written in *wnba_data_cleaner.R*.

In *wnba_data_cleaner2.R*, we take the data frame generated by *wnba_data_cleaner.R* and pivot the data frame wider so that the performance metrics each get their own column. The columns are then converted to variable type numeric as *pivot_wider()* keeps the values as type character. New variables called field goals percentage, three point field goals percentage, free throws made percentage, rest days between games, and performance metrics per 40 minutes (points, assists, rebounds, steals, blocks) are calculated. A treatment indicator variable indicating whether the game was back-to-back or non-back-to-back is also used to filter the data for players only playing back-to-back games.

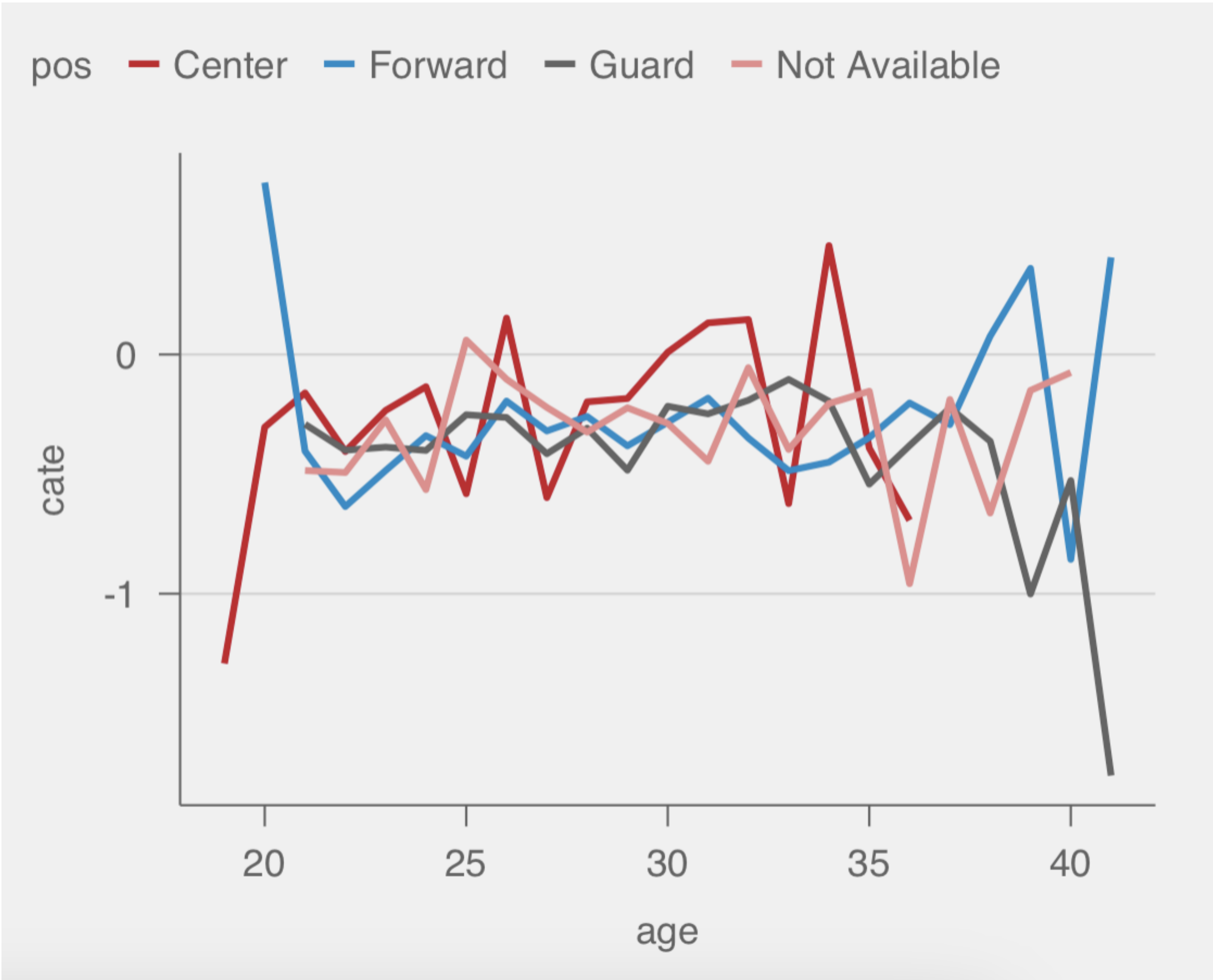
Ordinary Least Squares

In *2_1_OLS.R*, we load in our dataset cleaned by the two functions above, add splines to the age variable, and filter players within the age range of 19 to 38 with a sufficient number of occurrences. A set of features for analysis is defined, encompassing splines, team, player, and other relevant variables. Categorical variables are also converted to factors for categorical analysis. The script creates a grid representing different player attributes, teams, and seasons, setting the stage for subsequent modeling. The core of the analysis utilizes a for-loop over our target variables, such as the per 40 minutes metrics and field goal percentage, and seasons. In the loop, data is randomly sampled for training and an ordinary least squares (OLS) model is trained using an external sourced script. Predictions are generated for each combination of players, ages, and treatment conditions across seasons. The script concludes with a visualization of the results – a chosen metric and season are read from models generated from the loop saved in the "output/model" directory and a line plot depicting the age and predicted outcome is visualized. From the spline regression example plot below, we can see that for assists per 40 minutes for the 2012 season, assist outcomes generally increase as age increases. Reasons for this might be because senior players might make less risky decisions leading to an increase in assist opportunities or veteran players may experience a decline in athleticism, so to compensate they might focus on aspects of their game that rely on finesse and playmaking.

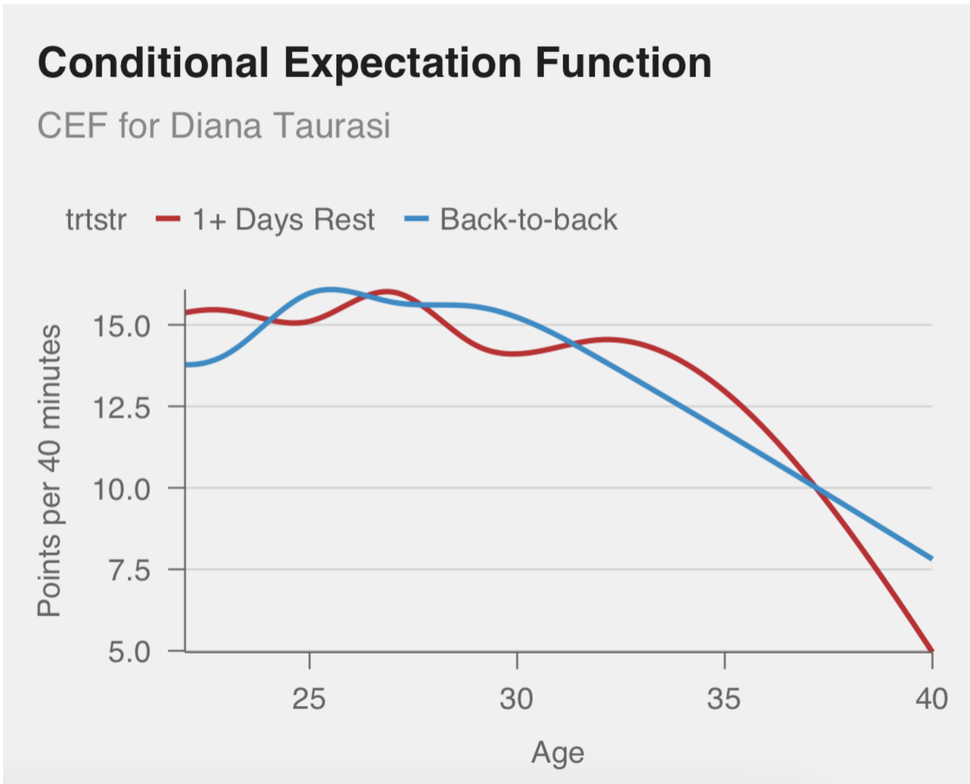


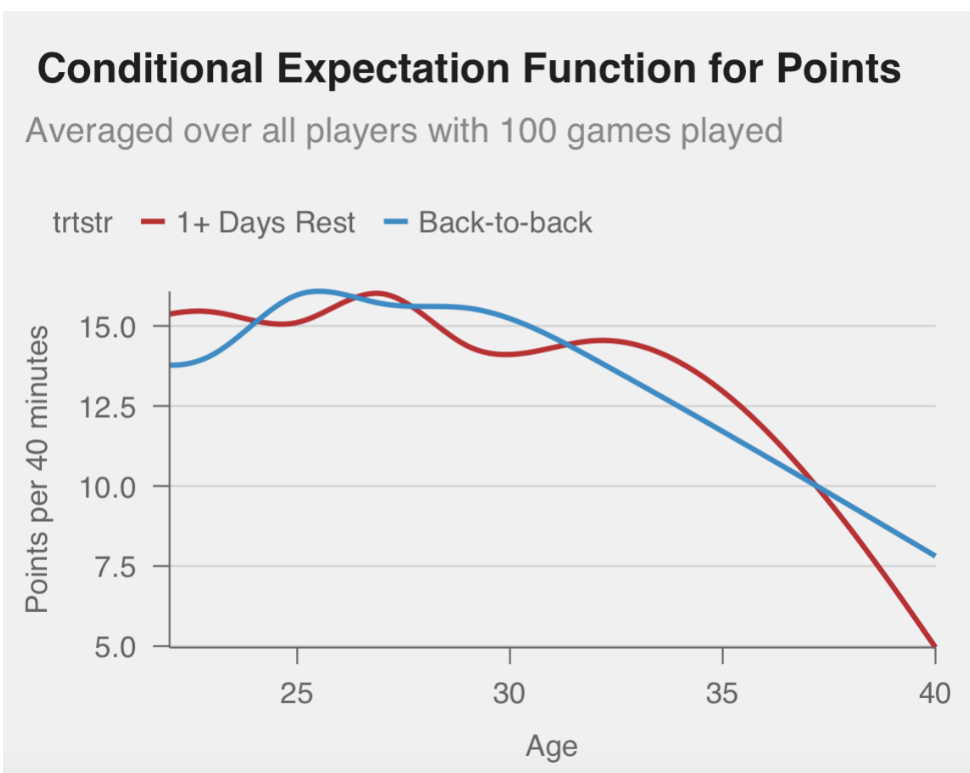
Meta Learners

2_Meta_Learners.R uses the OLS code, non-linear splines, and machine learning techniques to conduct an analysis of basketball player performance metrics, focusing on the impact of rest days on a player's scoring efficiency. For the data preprocessing, we follow the same methodology as *2_1_OLS.R*. The code visualizes CATE using a line plot illustrating how to CATE varies with age and player position, which is shown below. From the plot we can see that the curves jump up and down for each position throughout the entire age range, so this is probably not super helpful.



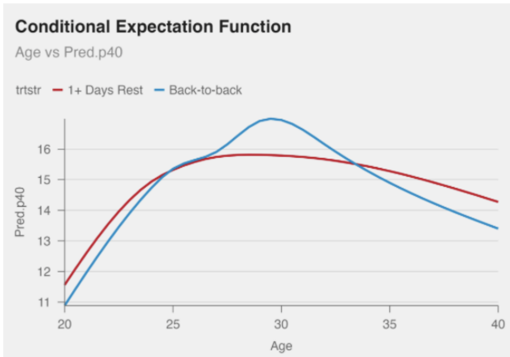
We also conduct a player-specific analysis and construct a curve showing how the predicted performance changes with age for players with varying rest conditions. Below we use the WNBA GOAT Diana Taurasi as our prime example and construct the conditional expectation function (CEF) curves for 1+ days rest and back-to-back games played. We can see that from age 20 to 30, her points per 40 minutes went back and forth for rested and back-to-back but for the majority of her thirties her output declined when she played back-to-back games versus when she had rest days between games. We then constructed a curve for the CEF for points per 40 minutes averaged over all players with 100 games played and we see that the curve follows the same pattern.

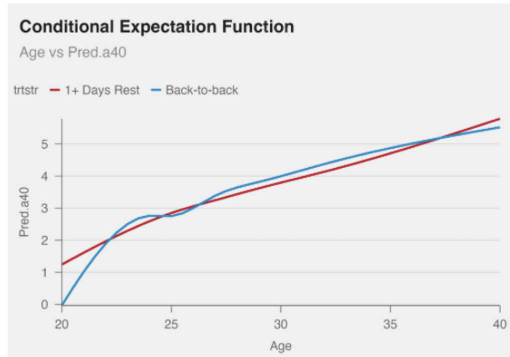


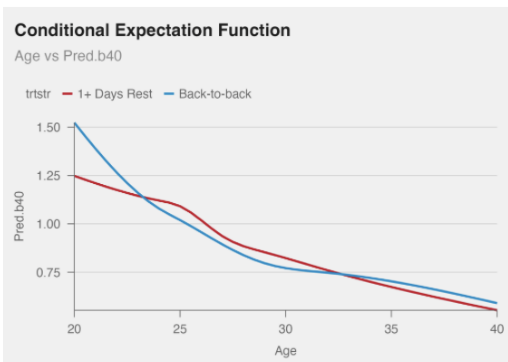


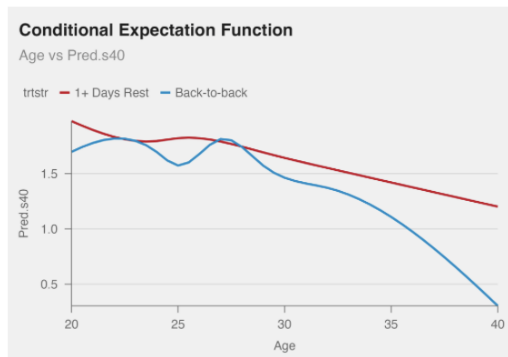
Conditional Expectation Functions

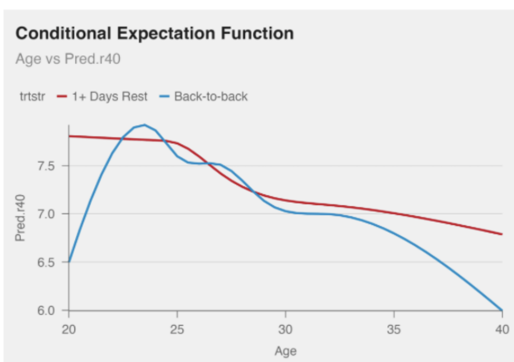
3_CEF.R takes the methodology we used in *2_Meta_Learners.R* and constructs CEF curves for the 1+ days rest treatment and back-to-back and creates plots for the predicted 40 minute metrics points, assists, rebounds, steals, blocks, field goals percentage, and free throws made percentage. The Age vs Pred.p40 curve shows that point output peaks on the back-to-back curve at the 30 mark then declines, going below the 1+ days rest curve. This could potentially be because basketball players tend to peak at around 28-29 years old and having the momentum of playing back-to-back games could be why they are in form at this age. Assists for both curves on the Age vs Pred.a40 plot consistently increase with not much difference between the two besides the lower endpoint where assists are higher in an athlete's early twenties when having rest between games. Blocks decline for both curves with the athletes playing back-to-back games showing a steeper decline. Steals show a similar trend where athletes show a steady decline in steals when rested, but a steep decline when playing back-to-back games – this makes sense because defensive responsibilities are physically more demanding than offensive ones and a younger player would probably have more energy and sharpness when making steals. Rebounds show an initial steep increase in the early twenties then a steep decline from the mid-twenties for back-to-back games. For rested games, we see a sharp decline from 25 to 28. The Age vs Pred.field_goals_pct plot displays a nearly linear increase in the field goals made percentage with the back-to-back curve under the first curve for most of the plot, meaning that the field goals made percentage is higher when athletes have rest between games. The Age vs Pred.free_throws_made_pct plot shows an interesting insight – the free throws made percentage is higher for back-to-back games for all ages. Players thrive on maintaining a consistent rhythm and playing a higher volume of games could allow them to stay in the flow state. Back-to-back games could keep players mentally sharp and more attuned to their shooting technique behind the free-throw line.

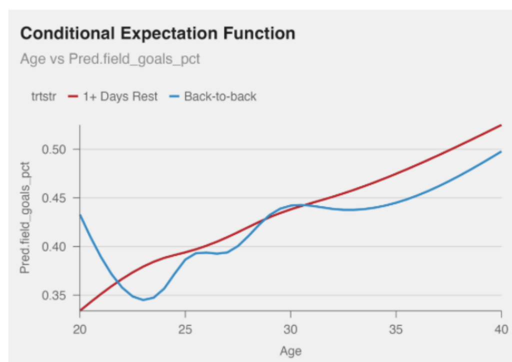


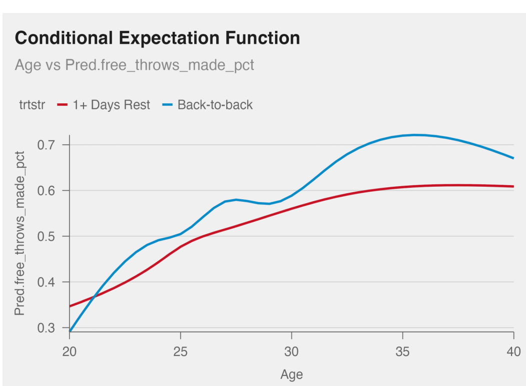




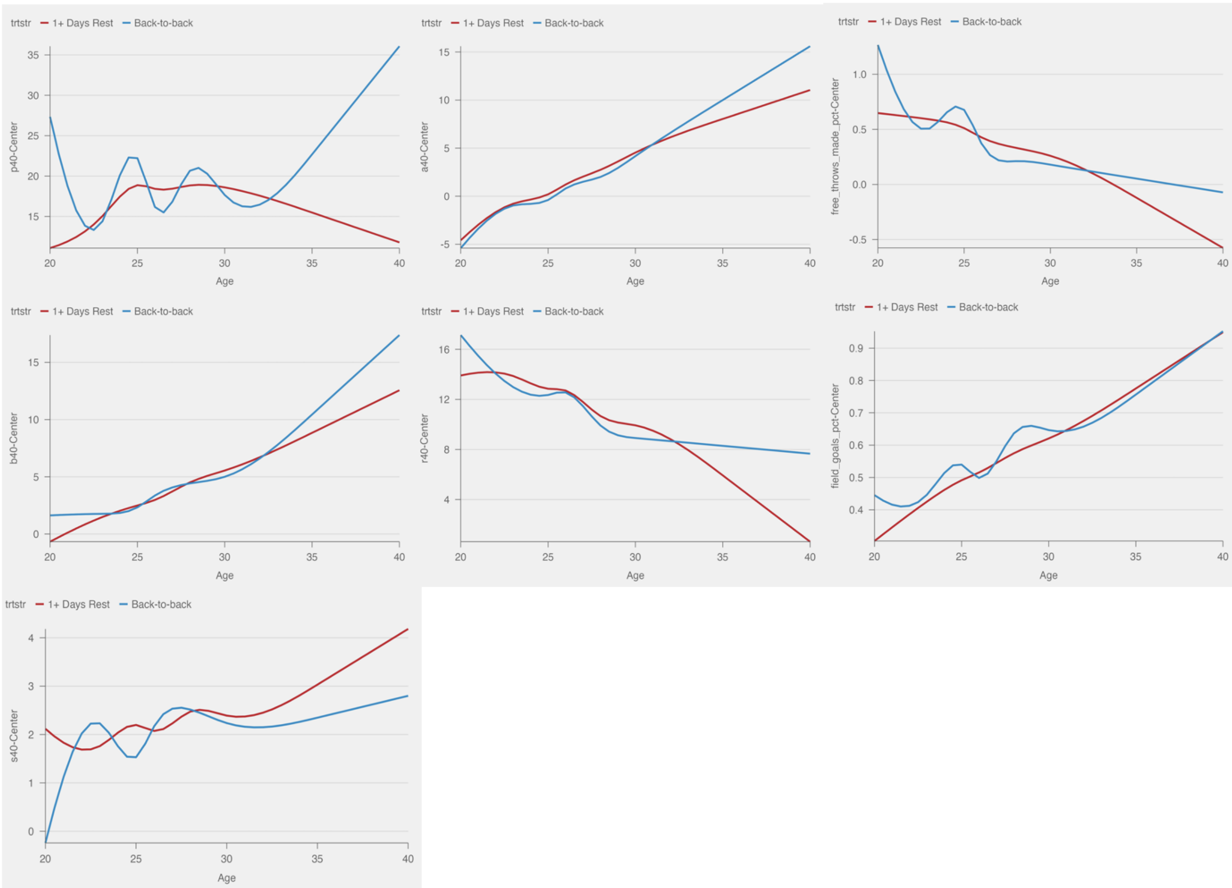




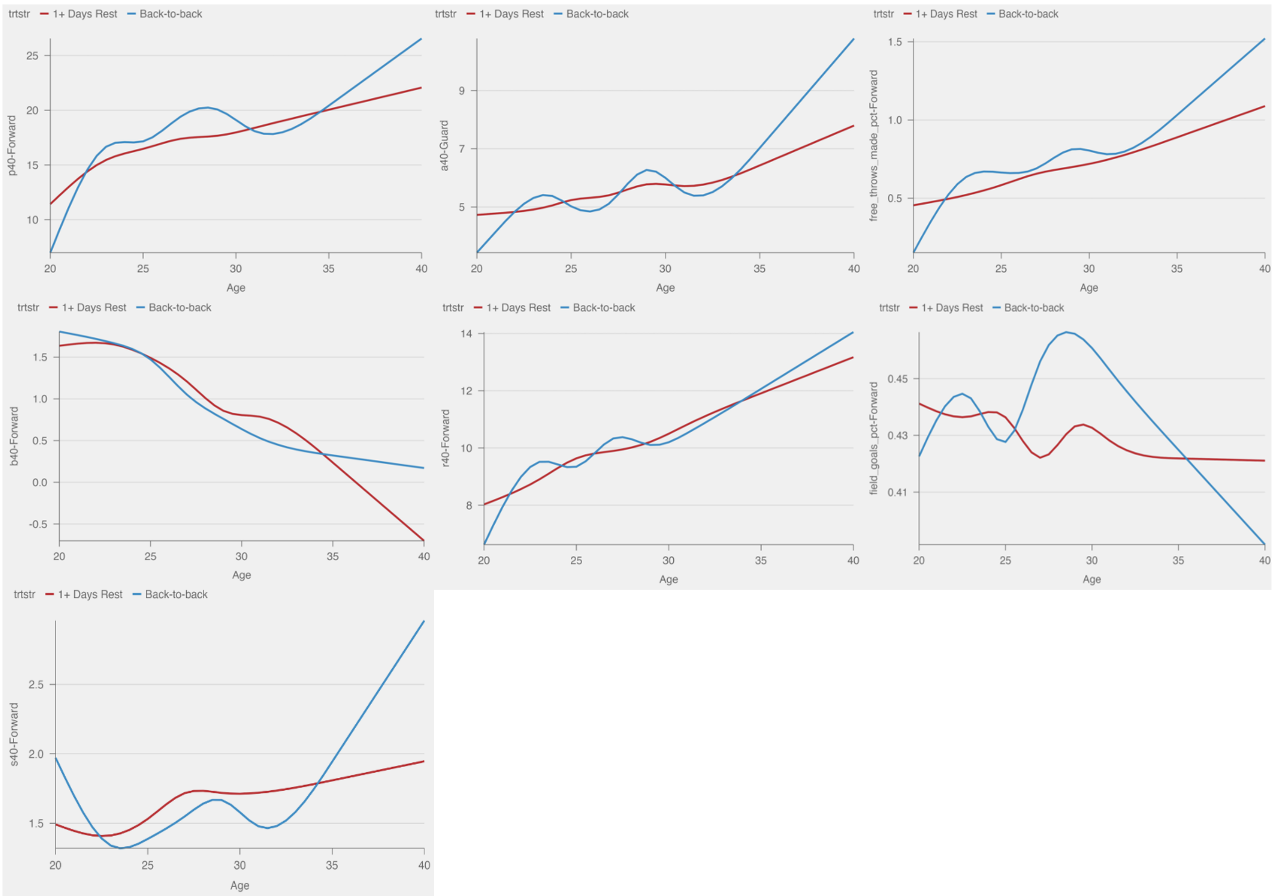




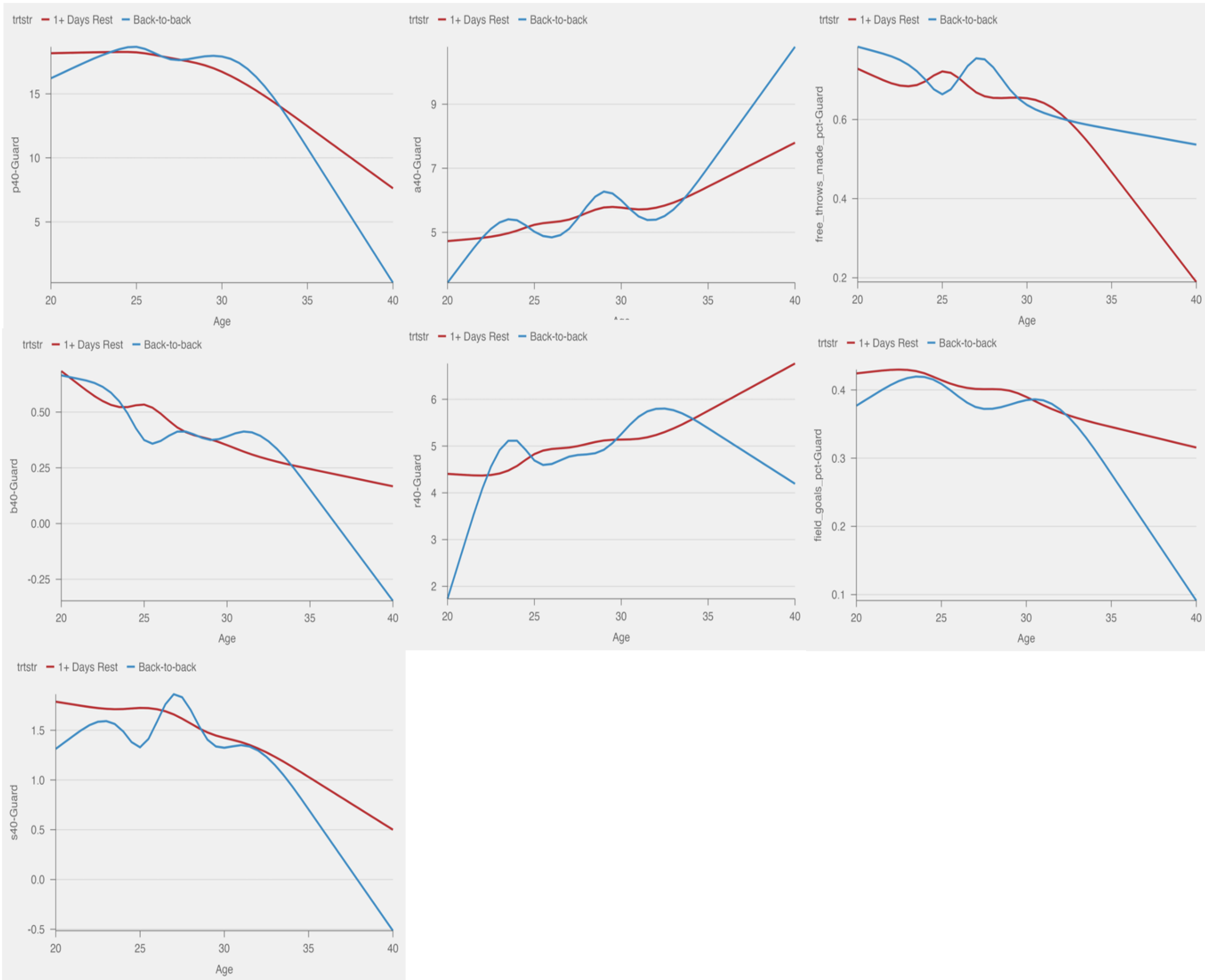
CEF Plots for Centers



CEF Plots for Forwards



CEF Plots for Guards

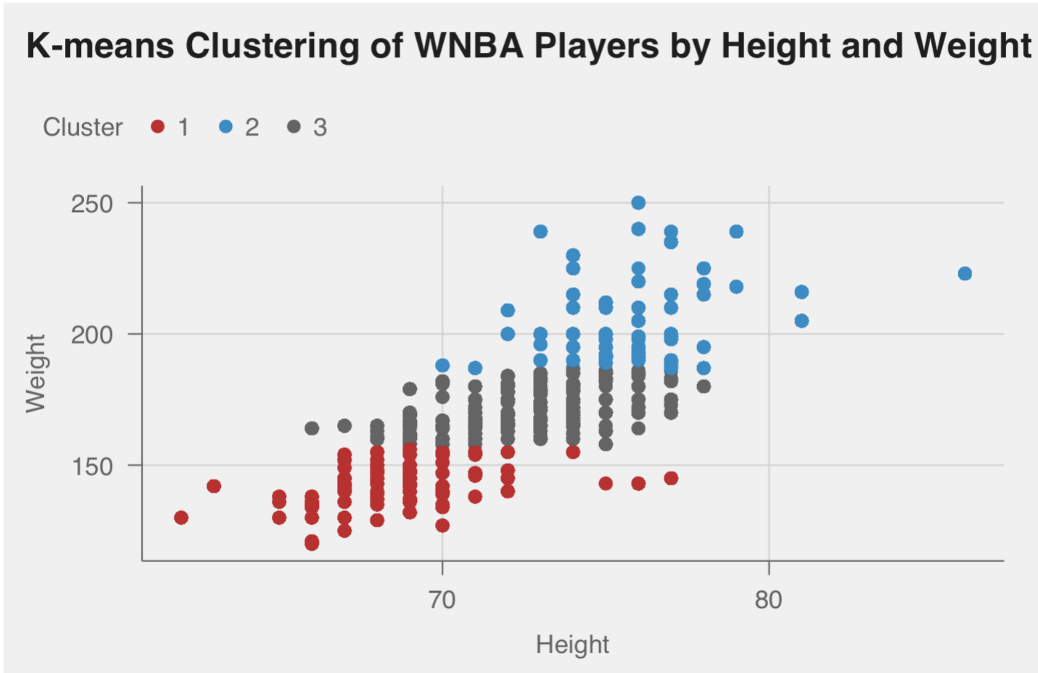


Plotted above are the CEF plots displaying the age-conditioned treatment curves for 1+ days rested games and back-to-back games by positions center, guard, and forward with points, assists, steals, blocks, rebounds, field goals made percentage, and free throws made percentage as the outcome variables. The curves mostly show what we can already expect, but there are some interesting points of conversation. Most of the 1+ days rest curves show more consistent, gradual lines while the back-to-back curves show multimodality and erraticity. The Age vs Predicted Points per 40 Minutes curve for centers show a multimodal curve for the back-to-back curve while the 1+ days rest curve shows a steady incline then decline. The two curves show a noticeable disparity starting in the 30s, with points per 40 minutes for back-to-back games increasing with age and decreasing with age for 1+ days rest. Additionally, free throws made percentage for all positions seem to be higher when playing back-to-back games for older ages – the disparity between the lines is the biggest for guards.

Biometric Clustering

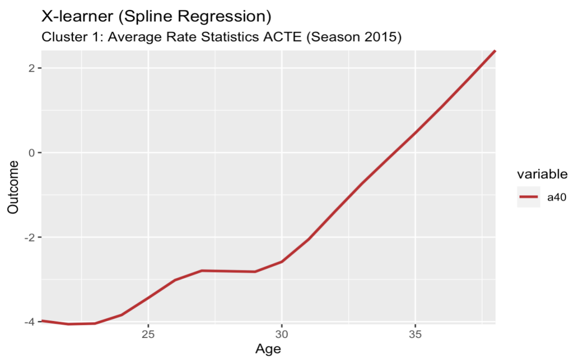
We performed biometric clustering to unveil more insights into the intricate relationship between WNBA player performance and the demanding schedule of back-to-back games. This was done by categorizing players based on their biometric attributes, specifically height and weight. Our aim was to discern nuanced patterns within distinct player groups to recognize that athletes with similar physical characteristics may exhibit unique responses to the challenges posed by consecutive games.

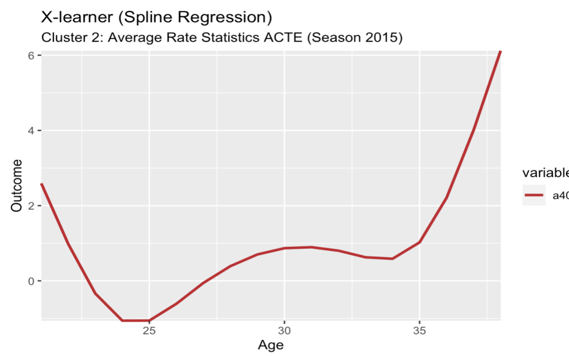
The incorporation of biometric clustering adds depth to the analysis. By grouping players with similar physical characteristics, the study implicitly acknowledges that individual player responses to the demands of back-to-back games may be influenced by unique biometric profiles. This approach allows for a more granular examination, recognizing that diverse player cohorts may exhibit distinct performance trends. Hence, we used the helper script *2_1_biometric_clustering.R* in order to cluster our data into three distinct groups based on height and weight.

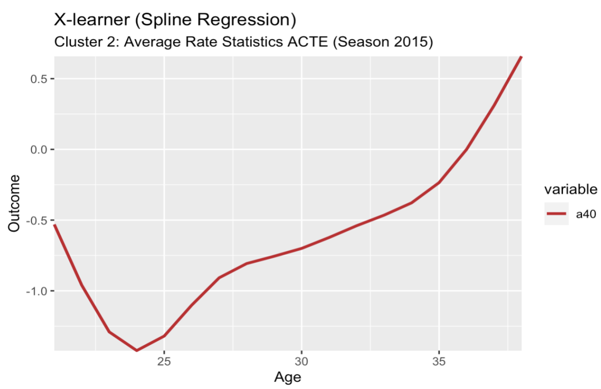


2_1_OLS_Clustered.R deals with modeling the clustered data in three distinct groups. It is similar to the *2_1_OLS.R* script, modified for biometric clustering. Once the data is loaded, it is clustered into three groups as shown in the plot above, and then splines are added to the age variable, and players are filtered within the age range of 19 to 38 with a sufficient number of occurrences. Thereafter, it follows the same modeling flow as used for the non-clusterd complete dataset.

From the spline regression example plot below, we can see that for assists per 40 minutes for the 2015 season, even though assist outcomes generally increase as age increases, the curve does not look the same for the three groups of clustered players. A general increase is expected as senior players might make less risky decisions leading to an increase in assist opportunities or veteran players may experience a decline in athleticism, so to compensate they might focus on aspects of their game that rely on finesse and playmaking. However, it is intriguing to see that cluster 2 (the cluster with the highest height and weight) is the one with the highest increase in assists. This could be because taller and heavier players in cluster 2 might predominantly play specific positions, such as forward or center, where playmaking responsibilities often increase with experience. As these players age, they may transition into roles that involve more facilitation of offensive plays, contributing to the observed rise in assists.







Shiny App

Link: https://sahil-singh-6.shinyapps.io/ACTE_WNBA_App/

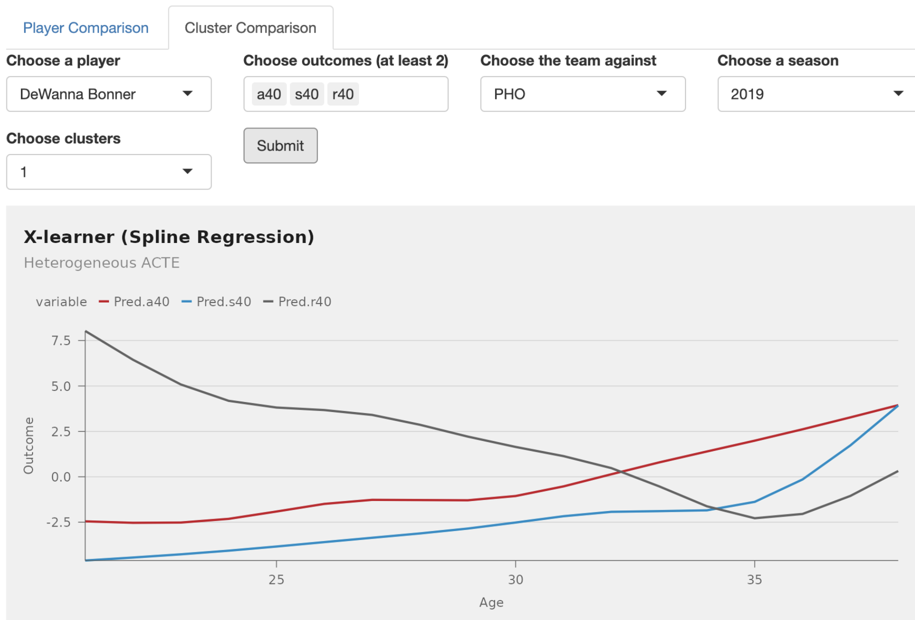
In order to create an user-friendly way to analyze age-curves, we have built a shiny app. This app aims to provide an interactive platform for comparing and visualizing the performance of WNBA players. The app is divided into two main sections:

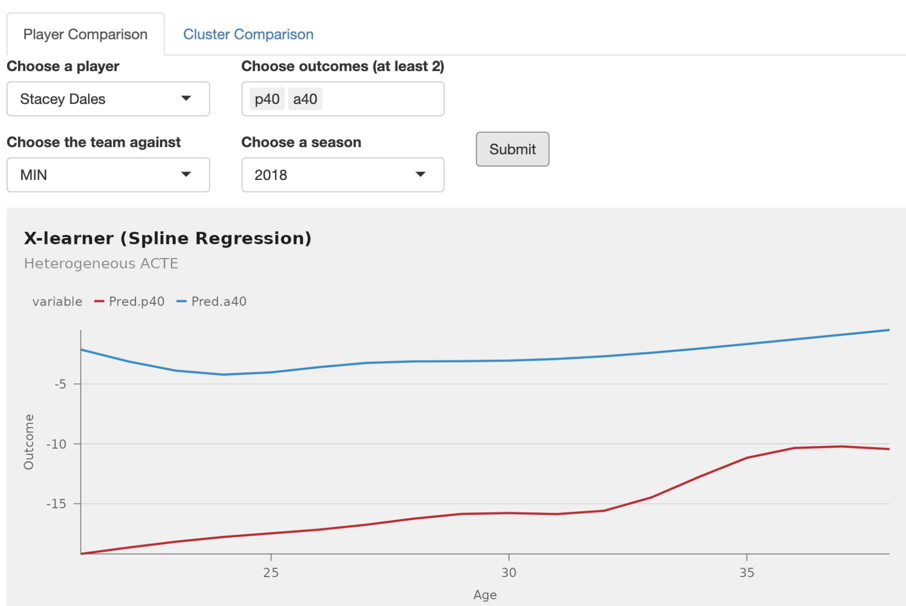
Player Comparison: Here, users can choose a specific player, outcomes (performance metrics), the opposing team, and the season of interest. Upon clicking the "Submit" button, the app generates a dynamic line plot illustrating how the selected player's predicted performance outcomes (such as points, assists, rebounds, etc.) vary with age. The plot employs an X-learner approach with spline regression, offering insights into how different metrics change over a player's career.

Cluster Comparison: Here, users can choose a player, outcomes, opposing team, season, and a specific cluster number. Clusters are groups of players based on their biometric clustering on weight and height. Upon clicking the "Submit" button, the app generates a line plot similar to the Player Comparison section. However, this time, it focuses on a specific cluster, allowing users to compare the predicted performance outcomes of players within that cluster and see if it differs for players in general for other clusters too.

In summary, this Shiny app provides a user-friendly interface for exploring and comparing the predicted performance of WNBA players, considering different metrics and the influence of player clusters. It offers an engaging and informative visual experience for users interested in understanding how player performance evolves over their careers.

The number of available seasons has been limited to three (2018, 2019, 2020) since shiny was not allowing the large number of data files (close to 5 GB) to be uploaded which would be required to include all seasons from 2003 to 2022.





Part II: Bootstrapping for Error Estimation of ACTE

Bootstrapping Motivation

The current model of age-conditioned treatment effects aims to find statistical significance in the difference in performance of players between back-to-back and non back-to-back games as a function of age. Currently, our model currently uses ordinary least squares regression (OLS) on covariates age, team, player, team-against, home-indicator, and season as our covariates to account for changes in response that may be due to these factors. In addition to these factors, teammates sharing the floor is also an important factor in determining one's statistical performance; yet, capturing the dependence in this is much more difficult that simply adding a covariate. Yet if successful, adding such a condition allows us to more precisely understand the effect of fatigue through back-to-back games, and allows us to importantly create more accurate error estimates of our ACTE curves. Indeed, we do not currently have any reliable method for quantifying our confidence in the model, and adding error bars could help us provide context of the quality of our analysis.

Data

As we are building upon previously existing methodologies for ACTE, our data will be the preprocessed data available in the ACTE Github repository, specifically the file "1_Cleaned_df.rds." Briefly, this data contains major box score statistics such as points, rebounds, assists, blocks, steals, shooting percentages, and 3-point percentages, as well as advanced statistics such as offensive and defensive ratings, true shooting percentage, and the aforementioned counting major statistics scaled to per 100 possessions. For our linear model, the preprocessing also includes other game-level data such team-against, home/away, and important an indicator for if the game was back-to-back which is our treatment that we are controlling for.

While our ACTE has the potential to predict for all box statistics using the aforementioned linear model predictors, we will focus on predicting points, rebounds, and assists categories. We do not demonstrate the ACTE prediction for predicting the other box scores due to computational limitations but recognize that doing so would only require a change of selected variables and use the same code. We will also utilize only the games information from the 2022 NBA season, and recognize that this simply necessitates a change of parameter in order to run on other years.

Bootstrapping Methodology

We refer to the dependency of performance between teammates as "teammate effects," and we choose to condition for teammate effects by utilizing one subsets of our game data; specifically instead of using all NBA player data available, we choose instead to only use stats of players when they are playing with another player for a significant amount of time. Our bootstrapping method can be outlined in the following steps:

- Create a list of all pairs of players that have spent significant "time together," and store the game ID of the games where these players played together.
- For each of n bootstrap samples:
    - sample a random pair, sample a random game which the pair played together, and store the box scores of the players,
    - repeat k times (k=size of bootstrap) to create one bootstrap

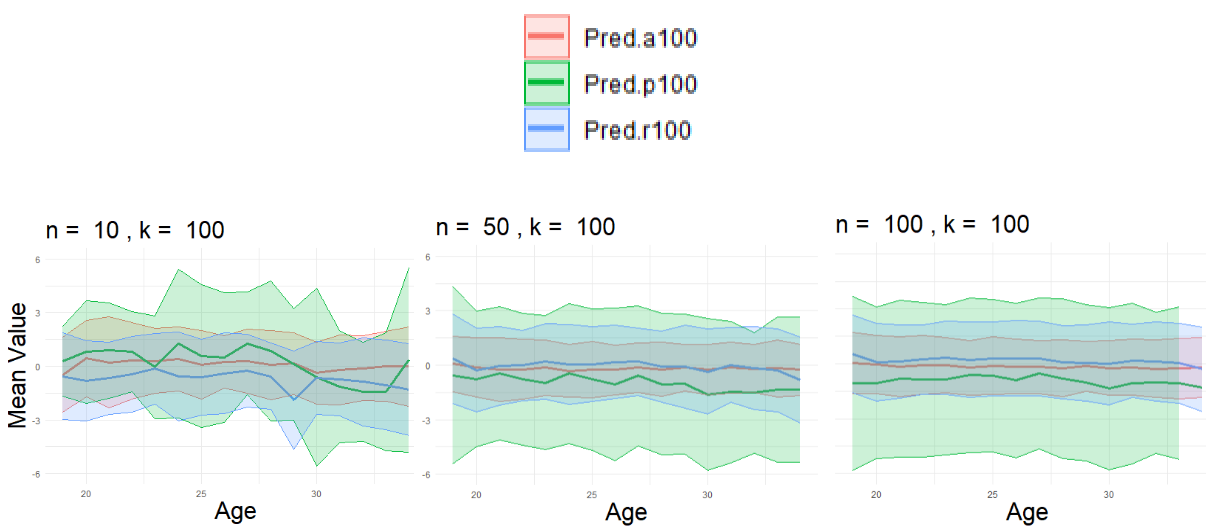
Note that in addition to our bootstrap variables, we also have a hypertunable parameter in how we quantify "played together." For our analysis, we proceed forward saying that two players "played together" if for a certain game, both players played more than 25 minutes. We recognize that the NBA does not provide a per-possession level of detail in their public information, and thus it is impossible to discern how many minutes or plays two players spent together in a game. In light of this, our method feels admissible since high player minutes for two players likely implies that the players spent a significant amount of time together on the floor in that game.

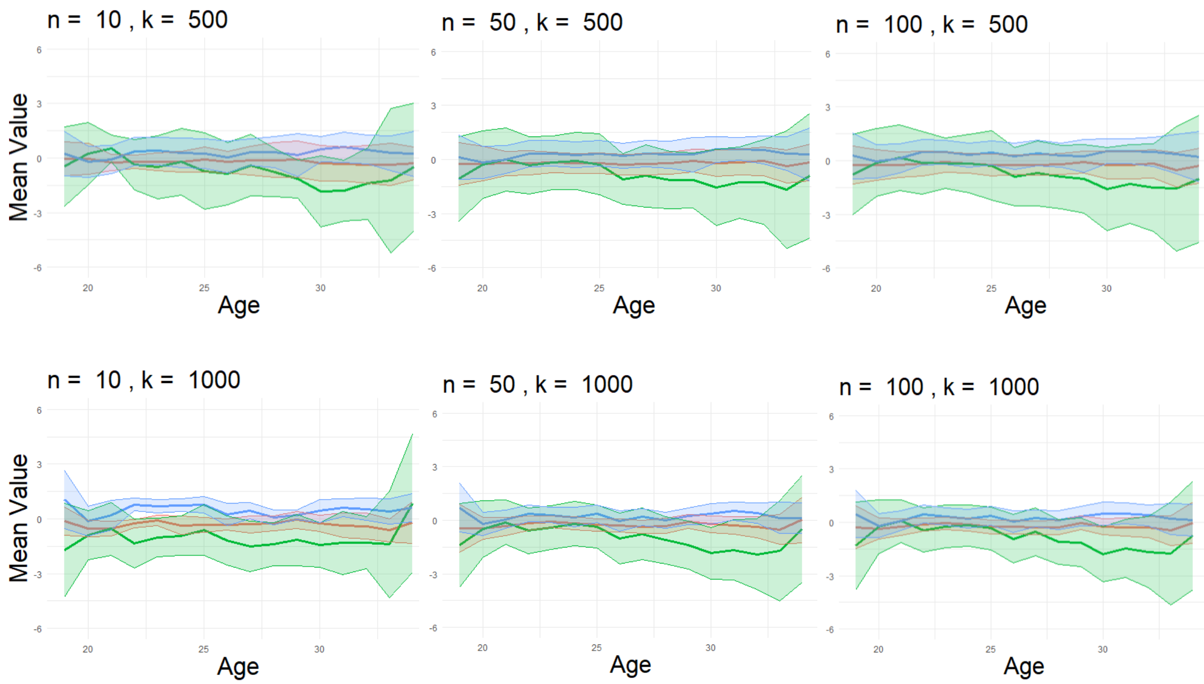
From here, we calculate the ACTE for each bootstrap sample to arrive at n estimates of the ACTE for each statistic of interest. We treat these n estimates as the distribution of our estimable parameters and use these standard deviations to plot the uncertainty of our models.

Bootstrap Results

Here, we present the graphical results of the ACTE with confidence bars for predicted assists, points, and rebounds per 100 possessions with varying bootstrap parameters. The bounds delineate one standard deviation distance from our mean, which approximately captures 68% of our bootstrap curves' predictions for each age.

ACTE Bootstrapping Results: n=10, 50, 100; k=100, 500, 1000

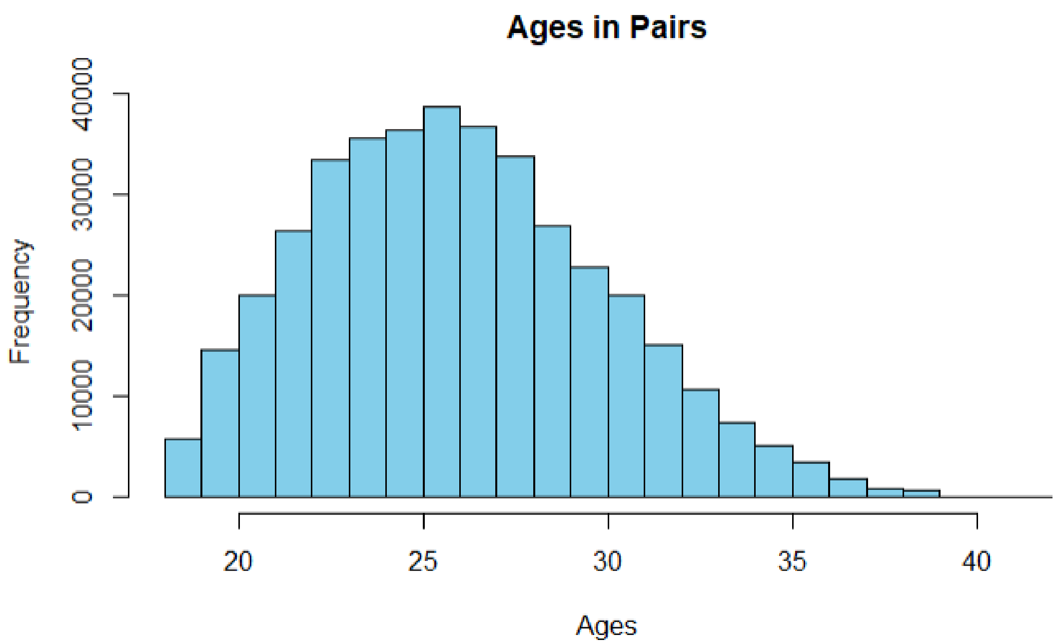




These plots vary the number of bootstrap samples between 10, 50, and 100, while also varying the number of samples per bootstrap between 100, 500, and 1000. By using the same graph scale, we can see the most obvious pattern that as the number of samples per bootstrap increases, the error bars of our ACTE also decreases. We also notice the same pattern per row-wise, whereas the number of bootstrap samples on error bars also decreased. This indeed corroborates with our existing literature that bootstrap confidence increases as both the number of samples and the number of bootstraps increase.

Across our statistics, our error bars seem to indicate that we have the least confidence in our ACTE estimates for points per 100 possession, compared to blocks and assists. This indeed makes sense because there is fundamentally larger variability in players' point scores, which reasonably varies between 0 and 30, while rebounds and assists per game tend to vary between 0 and 10. Thus, these large error bounds are not necessarily alarming.

As the confidence bands narrows as n and k increase, we seem to notice a pattern where confidence improves more noticeably for the middle ages compared to extreme young and old ages. We suspect that this is related to a data imbalance problem where we have more data on players in their mid-20s since younger players may still be inexperienced and not play many minutes, while older players have begun to decline in endurance. Indeed, our original categorization of pairs having "played together" was that both players played more than 25 minutes, and perhaps very old or young players are not represented as much in the pairs dataset with this information. To confirm this, we create a histogram of ages of players during games of our set of box scores relating to a pair, and see from the figure below that this is indeed the case.



Ultimately, the 1-sd confidence bounds we have set on these response variables all include 0, which seemingly indicates that there is no statistical significance of difference in performance according to these metrics between back-to-back and non back-to-back games at these ages. Indeed, this corroborates with our existing findings

Bootstrap Extensions

The aim of the bootstrap portion of this project was largely to create the code backend for further parameter tuning later in the larger project. As bootstrapping is a very resource-intensive process, we should pursue a more rigorous grid search of hyperparameters and their ACTE visualizations: currently, we would pursue a grid search of 100, 500, and 1000 bootstrap samples with 1000, 5000, 10000, and 20000 samples per bootstrap. There are also further investigations to be had about our definition of "played together," such as changing the threshold of minutes played needed between 10, 20, and 30 minutes per game. We also note that there are many different measures of success with respect to bootstrapping: while we do not explicitly define a measure of success for the purposes of this exercise, we recognize that bootstrap seeks to model the true sampling properties and thus may define one measure as the convergence in similarity between bootstrap runs using the same parameter. Towards this goal, we expect that the grid search with the largest parameters would yield the best results as long as we have enough compute power.

Besides the parameters, there are also different ways which we can modify our bootstrap methodology fundamentally. Though our current method involves sampling one game played together per pair of players, there exists an argument that batch effects are not adequately captured when only a one-game sample is taken; instead we could opt to select multiple or all games of each pair to theoretically capture all of the dependency. However, we do not pursue this here because it is difficult to conceptually justify and understand how sampling with replacement for bootstrapping might be affected when we select all the games for each sample (ie if we select a pair twice, then all of their games will show up twice, perhaps leading to too many duplicates within each bootstrap sample). Yet, we also recognize a current limitation of our model that pairs of players that have many games playing together are oversampled. As such, we preliminarily suggest a compromise methodology between the two that samples with replacement pairs, then also samples with replacement a constant number of games they've played.

Conclusion

In the complex realm of professional basketball, our study delves into the intricate connections between age, rest, and athletic performance. By leveraging rigorous methodologies, we analyzed age-conditioned treatment effects in the WNBA, shedding light on the nuanced variations in player metrics across different age groups and rest conditions. Bootstrapping for error estimation added a layer of depth, emphasizing the significance of understanding inherent uncertainties in our models. While our findings offer actionable intelligence for coaches and practitioners, it's crucial to recognize the multifaceted nature of sports, where data-driven insights intersect with unpredictable human dynamics. Our study not only enriches the discourse on load management but also underscores the importance of a holistic, data-informed approach to optimizing player performance and well-being in professional basketball.