



Maximizing Roster Efficiency in the MLS

Oliver Daboo, Otis Birnbaum
Colby College, Williams College
External Advisor: Daniel Wicker, Charlotte FC

Carnegie Mellon University
Statistics & Data Science

Introduction

Motivation

- MLS is the only major professional soccer league in the world with a salary cap, so it is extremely important to have the most efficient roster possible in terms of pay for each player and their performance on the field

Main Question

- How should MLS teams construct their roster and salaries based on a players position, age, and nationality

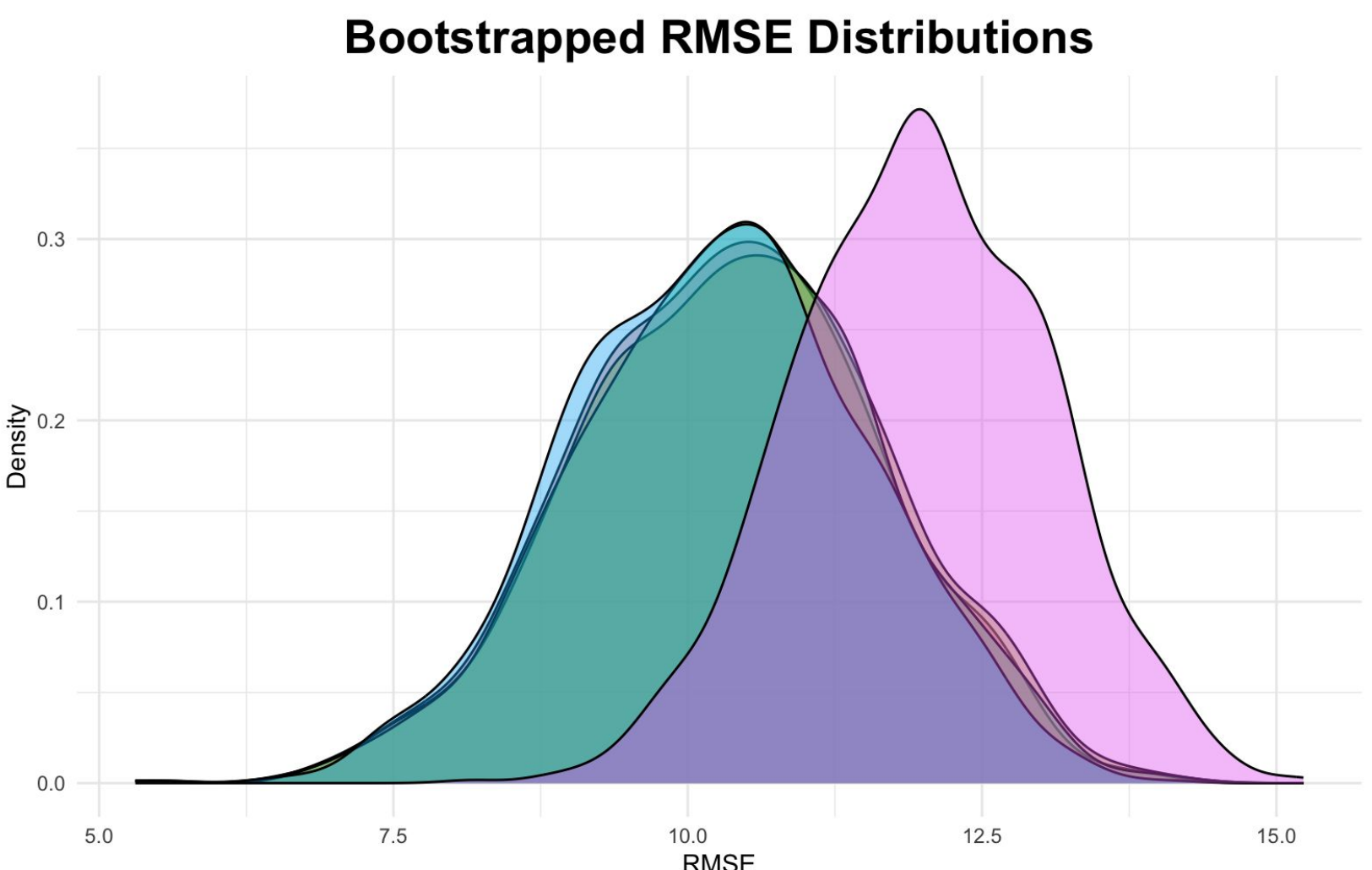
Data

- MLS Seasons 2021-2024 from **American Soccer Analysis**
- Used xG Difference to evaluate team performance
- Used Goals Added, a value added metric, to evaluate player performance
- Created a goals added per 90 minutes played per \$10k (ga_per_90_per_10k) metric to find undervalued players
- Used Guaranteed Compensation instead of base salary for player and team salaries

Methods

Team Modeling

- Split up each teams roster into the percentage of their total salary they pay different slots of players
 - Split up into 6 groups of 3 of the top 18 paid players
- Predicted xG difference using linear regression, regularization models, and XGBoost
- Trained the models on data from 2021-2023, tested it on 2024 and used bootstrapped cross-validation to evaluate model performances



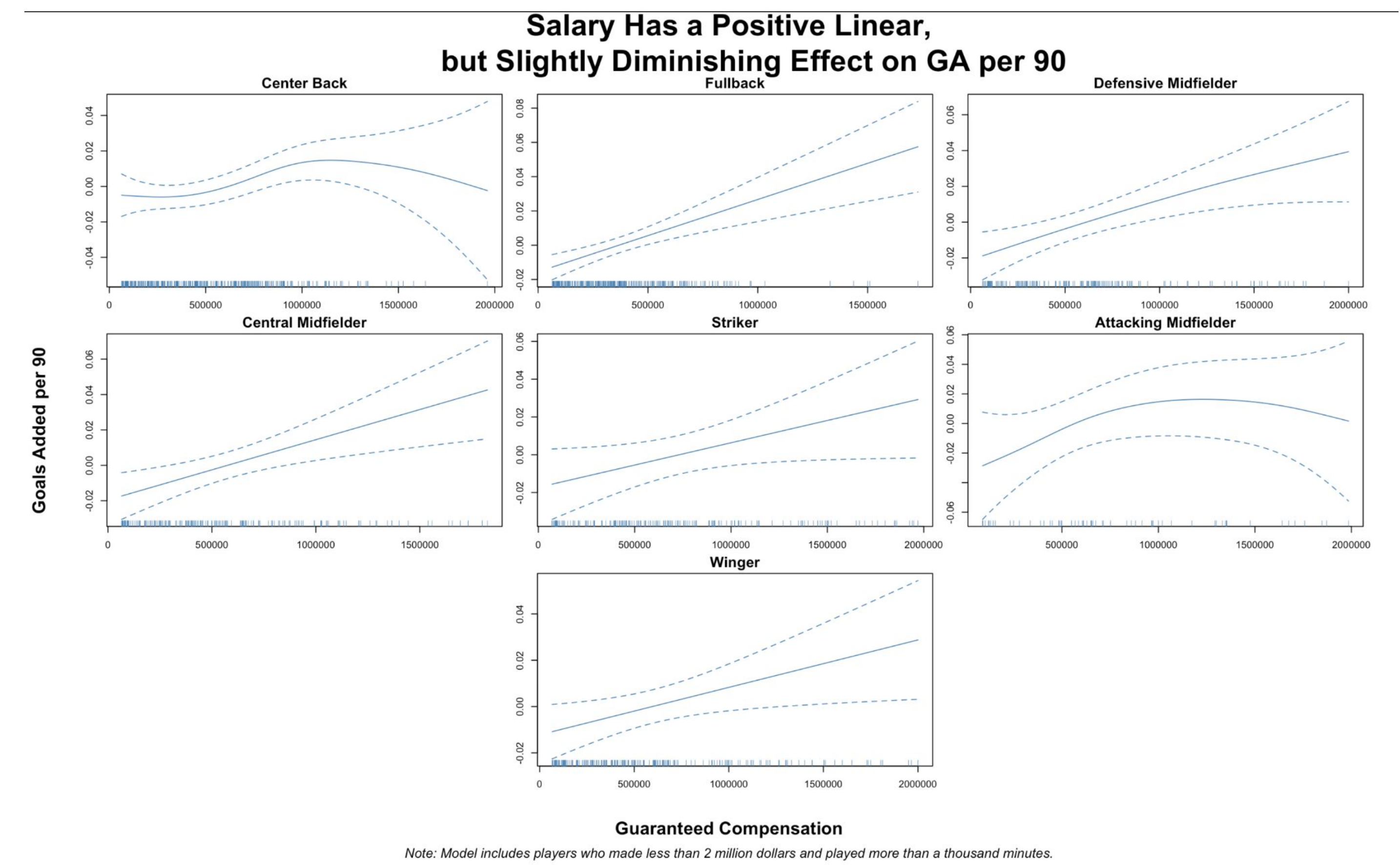
Player Modeling

- Fit a linear model to predict ga_per_90_per_10k based on player ages, positions, and nationalities
- Used Generalized Additive Models to identify non-linear relationships between salary and goals added among each position controlling for age and region.

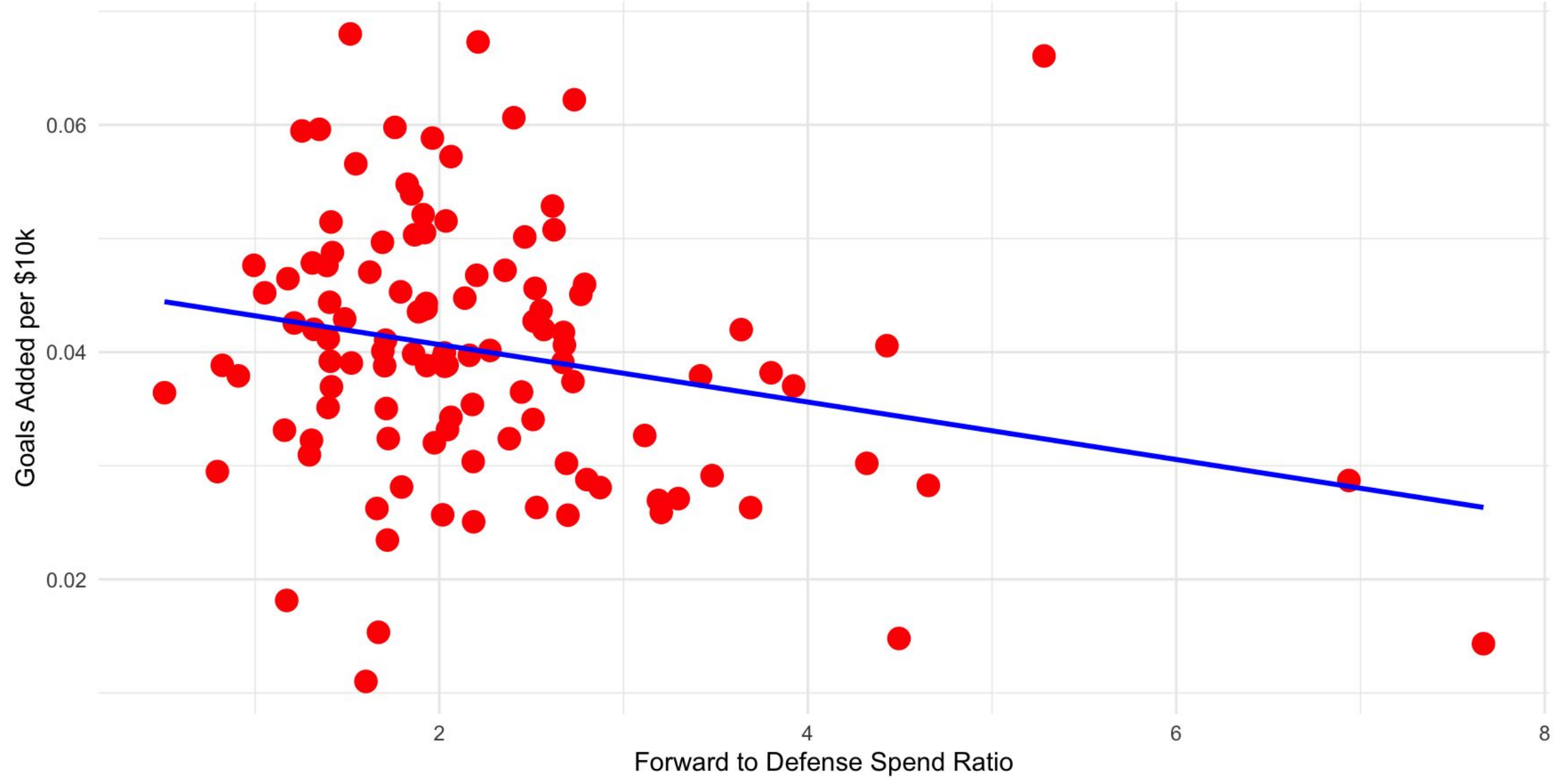
EDA and Results

The Top 5 Most Efficient Player Seasons in the MLS in the last 4 Years

Player, Year	Club	Position	Age	Nationality	GA_Per_90_Per_10k
Patrick Agyemang, 2024	Charlotte FC	ST	23	USA	.0475
Célio Pompeu, 2023	St. Louis City	W	23	Brazil	.0462
Tani Oluwaseyi, 2024	Minnesota United	ST	24	Canada	.0411
Jacob Murrell, 2024	D.C. United	ST	20	USA	.0369
Fredy Montero, 2021	Seattle Sounders	ST	34	Colombia	.0350

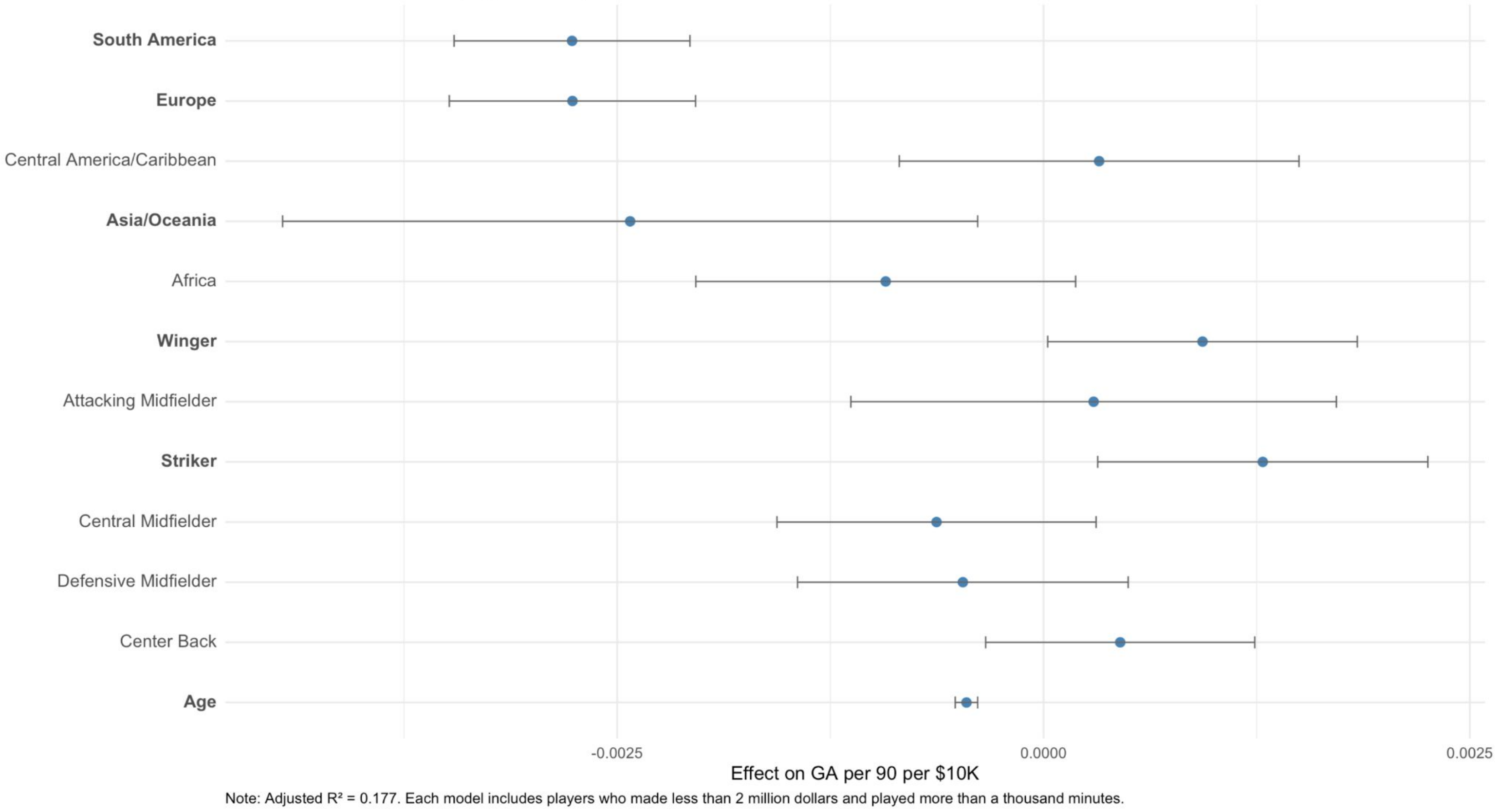


There is a significant relationship for the more that teams spend on their offense proportionally to their defense, the less efficient they are



Regional Origin, Positional, and Age Effects on Contract Value (Goals Added per 90 per \$10,000)

Bolded terms are statistically significant at $p < 0.05$. Benchmark Variables: Domestic and Fullback.



Discussion

Defense Wins Championships

- Teams should spend a higher proportion of their salary on defense than what is currently being seen in the league, especially if they have a low budget.
- (Mostly) Even Salary Spread**
- Teams should be spending a higher percentage of their current player salaries on the top 4-15 highest paid players, and a lower percentage on their top and bottom 3. This was determined by the Linear Model discussed on the left, that measured performance on salary spread.

Europeans, South Americans, and Older Players are Overvalued

- Being European or South American or being older all have negative effects on Goals Added per 90 minutes per 10,000 dollars spent.

Salary has Modest influence on Player Performance

- Broadly speaking, salary had a positive linear relationship with goals added per 90 across all positions, although some positions showed a diminishing relationship between salary and goals added per 90.

Acknowledgements

Thanks to Daniel Wicker, Dr. Ron Yurko, Quang Nguyen, and the entire CMSAC staff for providing support throughout the research process.