Impact of Turnovers in Early Season vs. Late Season Win-Rate In the WNBA

AUTHOR
Caroline Connolly, Pai Yang, Parth Harish, & Peyton Moffatt

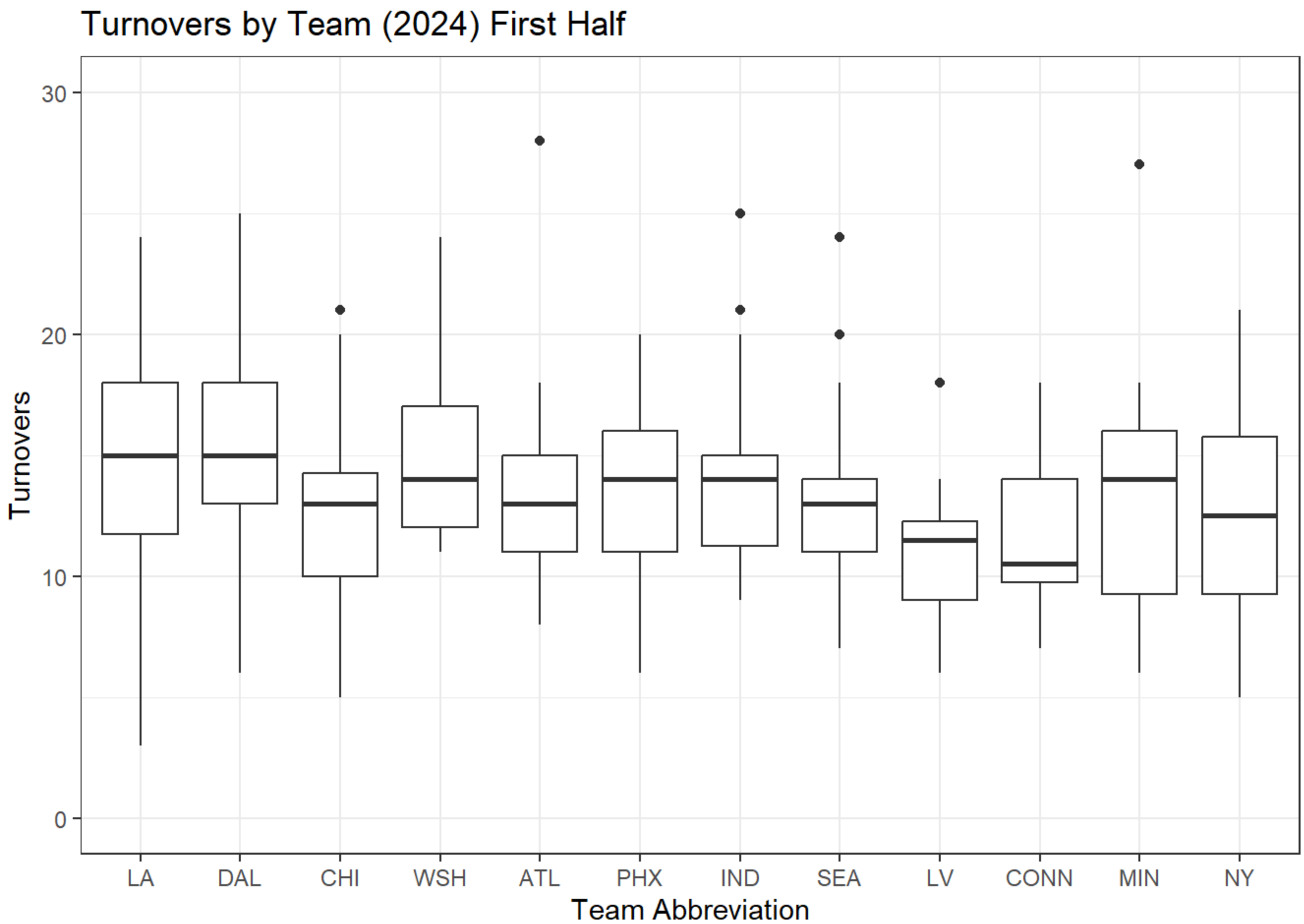
Introduction

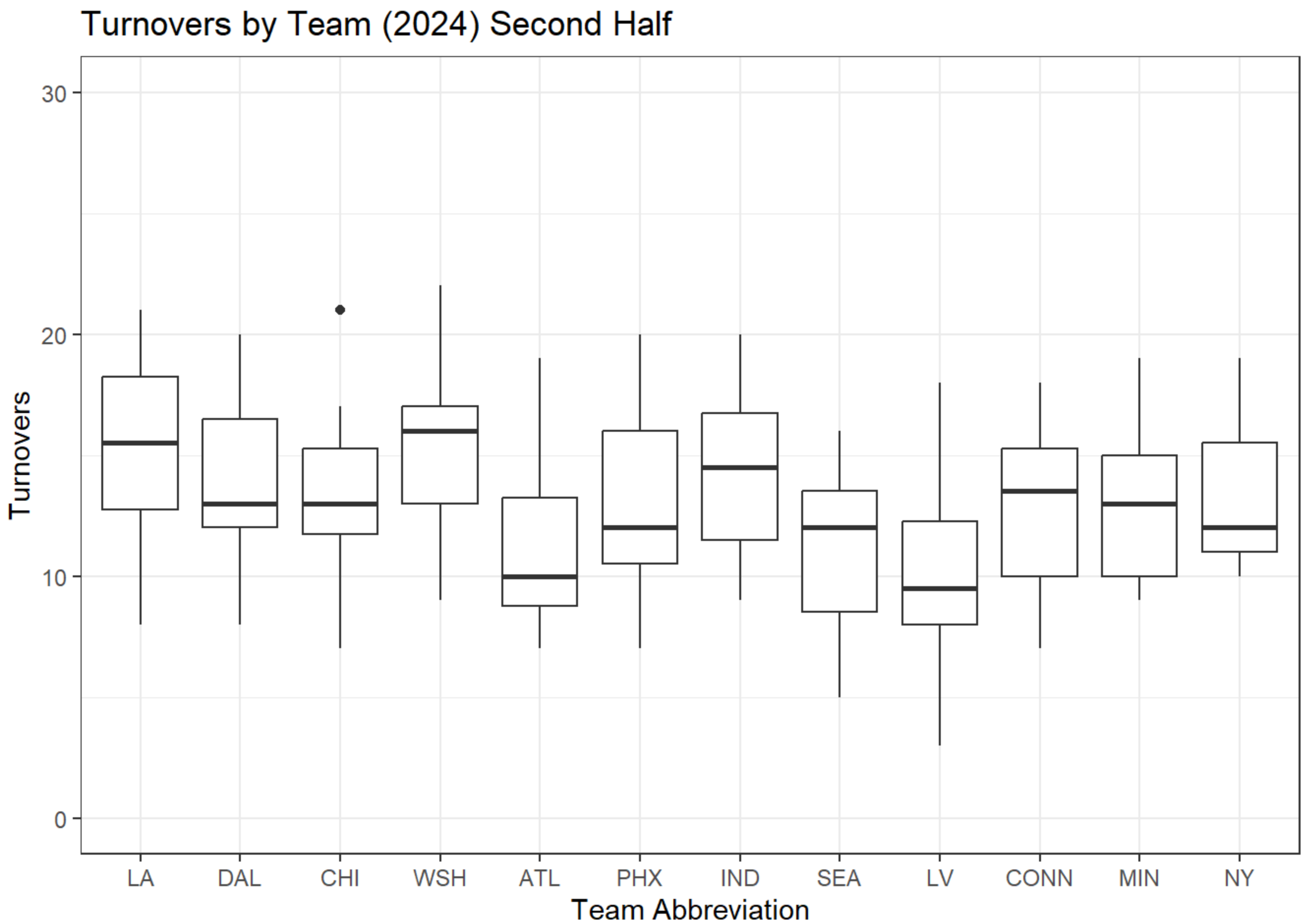
In this report, we are looking at the impact of early-season turnovers on late-season team performance in women's basketball, with a focus on the 2024 WNBA season. The motivation behind this comes from last season, where Caitlin Clark received a lot of criticism for her high turnover rates during the early part of the year. Despite this, her team's performance noticeably improved as the season went on. This leads to the question whether early mistakes—specifically turnovers—help contribute to better performance in late season games. Team with a large amount of turnovers are generally viewed negatively, but having a lot of turnovers at the beginning of the season could also be representative of a team adapting to new player or a new system. Thus, early-season turnovers could not just be mistakes, but an indicator of growth in the future. This idea is interesting as it challenges how we evaluate team performance, especially early on.. If early mistakes can help explain or even predict later success, that could shift how coaches, analysts, and fans approach the early part of a season.

To explore this, we used a generalized linear mixed-effects model and found that higher early-season turnovers were significantly associated with lower late-season win-rate across teams. This negative relationship was consistent across multiple methods, including bootstrap analysis and a generalized additive model.

Data

The data we used is the wehoop data set from the CRAN. This includes play by play and box score for both women's college basketball and WNBA, and we used turnover and win data for each team from the box score data for the 2024 season. We marked "early-season" as before the Olympic Break (May 14 to July 17) and "late-season" as after the Olympic Break (August 14 to September 19).


Turnovers by Team (2024) First Half
30
20
10
0
Turnovers
LA
DAL
CHI
WSH
ATL
PHX
IND
SEA
LV
CONN
MIN
NY
Team Abbreviation



The box plot shows team turnovers split by the first and second halves of the 2024 season, separated by the Olympic break as the midpoint. Teams are ordered by their overall regular season win-rate, with playoffs excluded.

One main insight is that teams with better overall records generally have fewer turnovers, and that teams having games with unusually high number of turnovers appear to be more common in the first half.

While both halves of the season show a negative relationship between turnovers and win-rate, first half turnovers appear to have more influence on win-rate in the graphs, suggesting that a team's ball control early in the season may have more impact on the overall team success.

Methods

For our project, we used two different methods of modeling: a generalized linear mixed-effects model (GLMM) and a generalized additive model (GAM), both assuming a binomial distribution for the binary response variable.

To start, we'll discuss the GLMM. Our GLMM is as follows:

```
glmer(is_win ~ early_season_turnovers + (1 | team_name),
      data=simple_box_scores_end, family="binomial")
```

We have `is_win` as a binary response variable indicating whether or not a team won, a predictor variable in `early_season_turnovers` which indicates a team's turnover per game in the first half of the season that in this scenario is a fixed effect for us, and e another predictor variable `team_name` which indicates the team's names and is used as a random effect. We use `team_name` as a random effect to give us a random intercept for each team, just for the simple fact that each team has a different baseline intercept. The model uses a binomial response and logit link function. We want to see if early season turnovers have an effect on winning games, and that's exactly what we are hoping to figure out from this model.

Next up is the GAM. Our GAM is as follows:

```
gam(is_win~s(early_season_turnovers),
    data=simple_box_scores_end, family="binomial")
```

Here we have the same binary response variable, `is_win`, and we again have our predictor variable of `early_season_turnovers`, but this time it is smoothed to allow for a non-linear relationship. This allows us to do a logistic regression without assuming a linear effect. and instead seeing a potential nonlinear trend.

We will compare these models first by seeing how well they each do at predicting a team's win/loss. We can easily do this by making predictions using each model and comparing the accuracy of predicted win/loss versus actual win/loss. We can also compare and see if they give us the same conclusion and whether said conclusion is significant/insignificant.

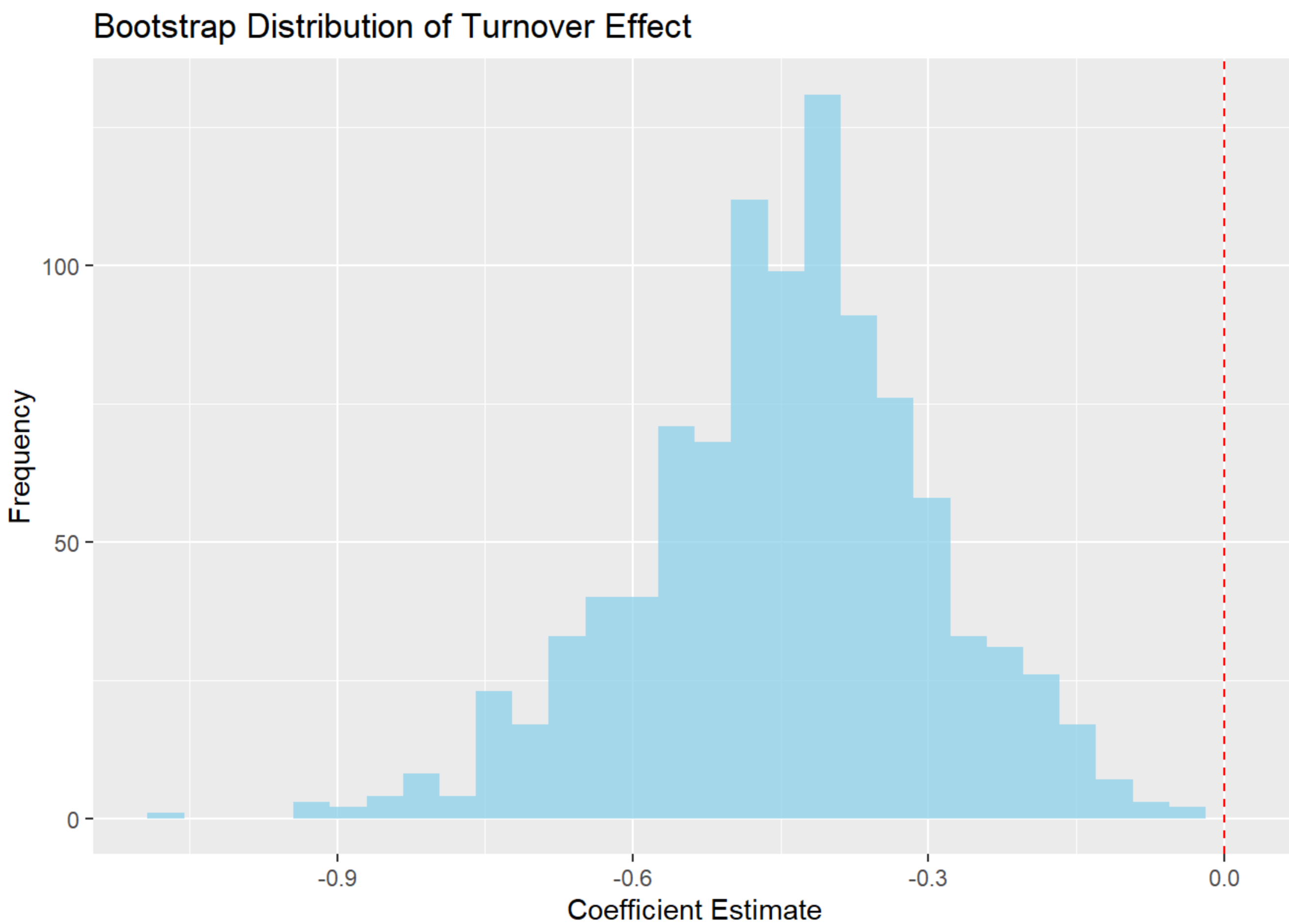
To quantify uncertainty, we will use case resampling bootstrapping for both models. For the GLMM, we will bootstrap the effect of `early_season_turnovers` as well as the intercept for `team_name`. We will also create 95% confidence intervals and see if we get significance or not for each. For the GAM, we will bootstrap and get confidence intervals at many values of `early_season_turnovers` and see how this compares. For both model, we are using 1000 bootstraps. Bootstrapping allows us to simulate many versions of each model to gather a bunch of data values and gives us a range. This is very helpful in checking for uncertainty and allows us to become a lot more certain of whatever conclusion we end up coming to. While it makes relatively few assumptions, one limitation is that sports data like the data we are using in our model often includes some dependence between observations. We try to account for this with team-level random effects in the model, but resampling the individual games may not fully account for the correlation between games within teams.

Results

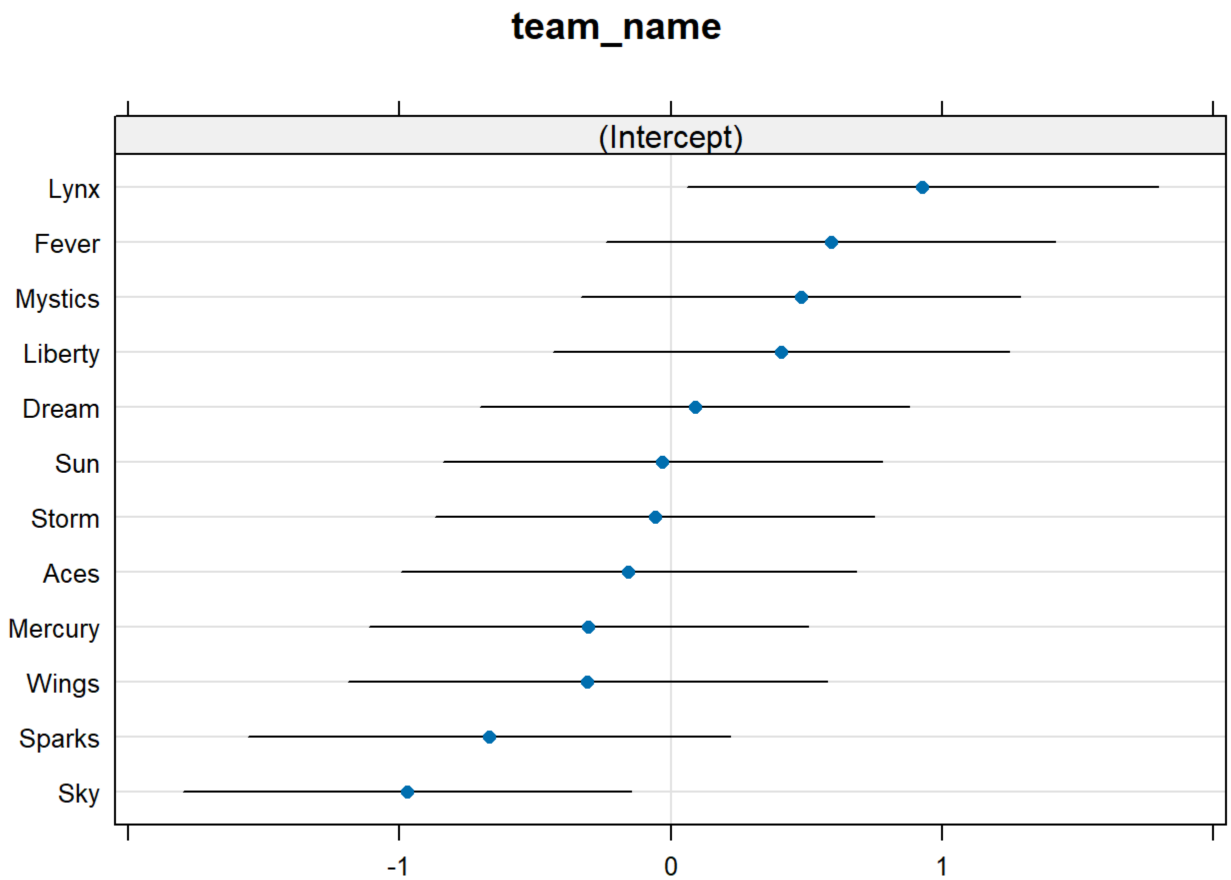
We used a generalized linear mixed-effects model (GLMM) to study the relationship between early-season turnovers and late-season win probability across WNBA teams. Early-season turnovers were treated as a fixed effect, and team-specific random intercepts were included to account for differences between teams.

From the GLMM output, the coefficient for early-season turnovers was estimated to be approximately -0.416 with a p-value of 0.0429, indicating that higher early-season turnover averages were significantly associated with lower odds of winning later in the season. This suggests that controlling turnovers earlier may positively impact team success later. The model's predictions correctly matched the actual win/loss outcome 67.93% of the time, offering more support for the relationship between early-season turnovers and late-season team performance.

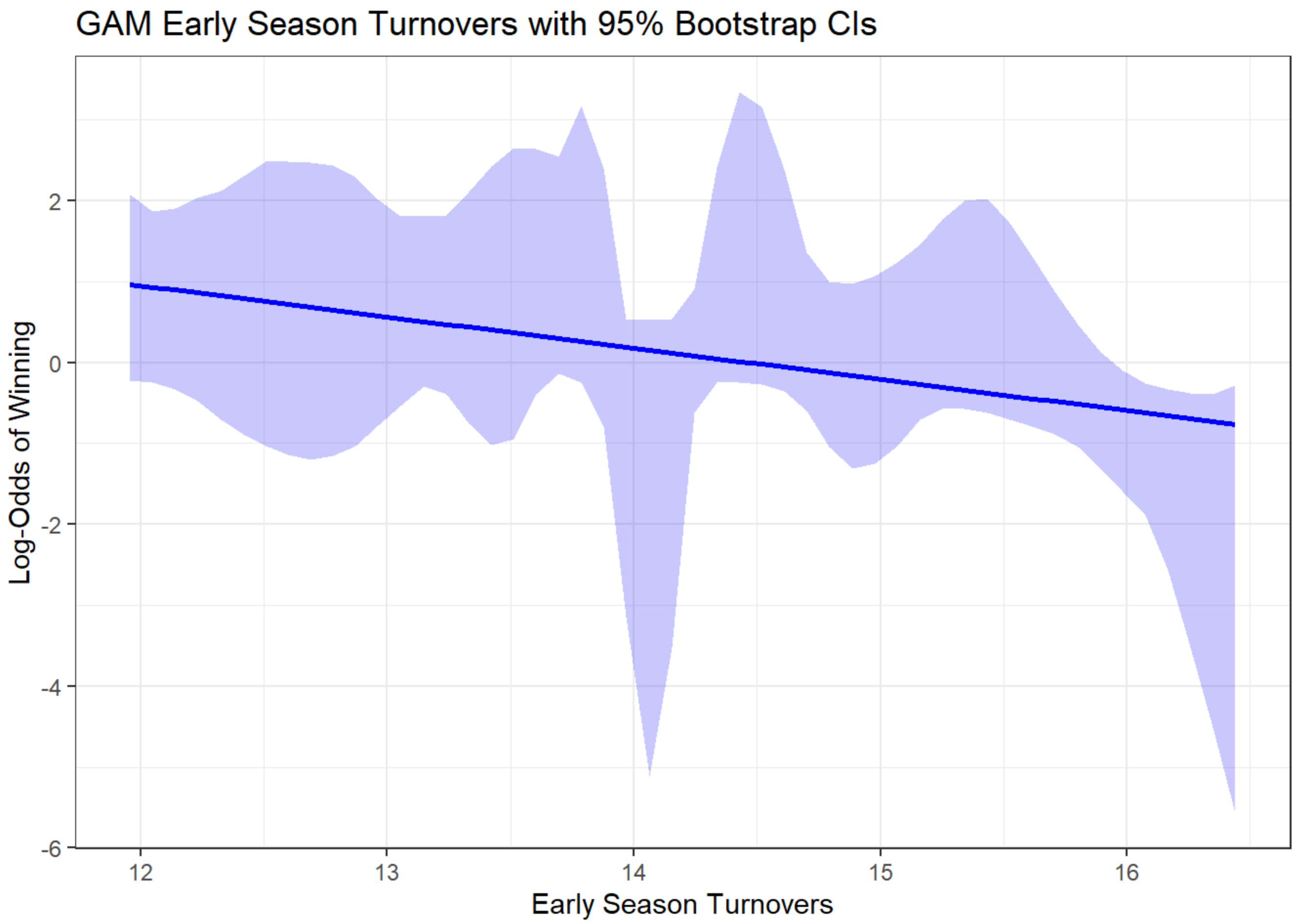
To further evaluate the model stability, we performed bootstrap analysis on the early-season turnover effect. The distribution of the bootstrap coefficient estimates was mostly negative and stayed well below zero, reinforcing the result that early turnovers negatively impact win probability.



We also visualized team-specific effects with a random intercept plot. While there was some variability across teams, most teams showed similar baseline performance patterns after controlling for early turnovers. No single team appeared to strongly deviate from the overall trend.



To check model robustness, we also fit a Generalized Additive Model (GAM) to allow for a potential non-linear relationship between early turnovers and win rate. The GAM confirmed a slight negative trend: as early-season turnovers increased, the log-odds of winning decreased. However, the relationship remained mostly linear and the explained variance was modest (Deviance explained = 3.95%), suggesting turnovers are an important but not sole predictor of later success. The GAM's predictions matched the actual win/loss outcome 58.70% of the time, which is lower than the GLMM's accuracy. This suggests that the GLMM's inclusion of team-level random effects improved the model's prediction accuracy, but since the increase is small, other unmeasured factors are likely at play.



Finally, bootstrap confidence intervals were calculated for both the model intercept and the turnover effect. For the turnover coefficient, the 95% confidence interval was approximately [-0.7571, -0.1604], which does not cross zero. This provides further evidence that early-season turnovers have a statistically significant negative impact on late-season winning chances.

Discussion

Our project looked at whether early-season turnover rates would be able to predict late-season success in the WNBA. We used a generalized linear mixed-effects model and a generalized additive model and found that higher early-season turnovers were associated with a statistically significant decrease in late-season win-rate. The GLMM estimated a negative turnover coefficient (-0.416) with a 95% bootstrap confidence interval excluding zero. GAM on the other hand confirmed a similar, mostly linear negative trend. When these results are put it together, they suggest that ball control early in the season may play an important role in setting the foundation for later team success.

However, there exist important limitations to our analysis. One thing that we didn't check was the correlation between early- and late-season turnover rates. If turnover rates are stable across the season, our model may simply reflect the broader link between high turnovers and poor performance, rather than identifying a developmental effect. In addition we modeled only team-level turnover averages and we missed game-level variability and player-specific effects particularly instability among rookies, who usually tend to have more volatile early-season performances.

Future work on this project could address these gaps by directly measuring turnover stability across the season, incorporating game-level turnover statistics, and distinguishing between rookies and veteran players to better capture learning effects. Expanding the analysis across multiple seasons could also help validate whether the patterns we observed in 2024 hold more broadly across the league.