NBA Defensive Impact

By: Joshua Gabriel Rusit

Abstract

This project examines the relationship between everyday defensive actions in NBA games and winning a single game. The data include one full season of team game box scores. I focus on five familiar stats that speak to defense. These are defensive rebounds, steals, blocks, turnovers, and personal fouls. I use a simple model that turns those numbers into a win chance. I also test the model on games it did not see before to make sure the results hold up. The main result is that getting more defensive rebounds and more steals is linked with winning more often. Giving the ball away and fouling are linked with winning less often. Blocks help too, but not as much as the other two. The story is clear, and it matches what coaches often teach. Ending possessions, taking care of the ball, and staying disciplined are key parts of winning.

Introduction

I wanted to study this topic because many people talk about offense first. Points and shooting get a lot of attention. But when you listen to players and coaches, they often say that defense is what wins tight games. I wanted to see if we could measure that idea in a simple way that anyone can read. I wanted a study that does not need advanced math to see what matters. So I focused on the parts of defense that appear in every box score. I looked at defensive rebounds, steals, blocks, turnovers, and fouls. Then I asked a basic question. When those numbers change, how does a team's chance of winning change?

What I Looked At

The data come from the 2023 to 2024 NBA season. Each row is one team in one game, with a win or a loss at the end. I used a model that learns the link between the five stats and winning. To make sure the results are not just a pattern in the training data, I held out part of the season and tested the model there. This helps show that the findings are real. It also lets me check if the predicted chances match what happens on the court. If the model says a team has a 60 percent chance to win, then teams like that should win about 60 percent of the time. That is an important check.

Methodology

To run this study, the first step was to load the season's box score file and clean it so every game had the same column names and formats. The win or loss flag was turned into numbers, where 1 means a win and 0 means a loss. The five defensive stats of interest were set to numeric so the model could read them. I also created a simple game identifier using the matchup text and date, and a basic home indicator using whether the matchup said "vs" or "@." These small steps made the data consistent and ready for modeling.

Next, the data were split into two parts. About eighty percent was used to train the model, and twenty percent was held out for testing. The split was stratified by wins and losses, so both sets had a similar balance. This matters because it lets the model learn on one set of games and then get checked on new games it did not see. It is a simple way to guard against overfitting and to get a fair read on how well the results hold up.

The main model was a logistic regression that predicts the chance of a win from defensive rebounds, steals, blocks, turnovers, and personal fouls. I also fit a second logistic model that included the home indicator to see if this easy context improves performance. To check stability, I fit a lasso version of the model. Lasso is a regularized method that can shrink noisy effects toward zero. I also tried two simple combined terms, turnovers plus fouls and steals plus blocks, to let the regularized model see basic interactions without making the model complicated.

After fitting the models on the training set, I used the test set to judge performance. I reported accuracy for a clear percentage of correct calls, a ROC curve and its AUC score to show how well the model separates winners from losers at many thresholds, a PR AUC to reflect precision and recall, and the Brier score to measure overall probability error. I also looked at calibration in two ways. First, I summarized games into probability groups and compared the average predicted chance to the actual win rate in each group. Second, I estimated a calibration intercept and slope, where an intercept near zero and a slope near one indicate honest probabilities.

For interpretation, I converted the logistic coefficients into odds ratios with 95% confidence intervals, allowing readers to see the direction and magnitude of each statistic's effect in plain terms. Because each real game appears twice in team form, I used cluster-robust standard errors by game to make the uncertainty estimates more realistic. This usually widens the intervals a bit, but it makes the results more trustworthy. Finally, I included simple diagnostic checks such as variance inflation factors, residual histograms, leverage and influence plots, and quick linearity checks on the model's logit. These checks help confirm that the model's assumptions are reasonable for this data and that no single point is driving the story.

Main Results and Why They Matter

The main finding is simple. Defensive rebounds and steals are strong signs of winning. When a team ends the opponent's possession with a rebound, that stops a second shot and often starts a new chance on offense. When a team makes a steal, it takes away a chance from the other team and creates a chance for itself, often in transition. Both actions raise the chance of winning. Blocks help too, but the effect is smaller. They can change shots and swing momentum, but they are less consistent than rebounds and steals. On the other side, turnovers and fouls work against you. A turnover gives the ball away for free. A foul can give up free throws or put the other team in the bonus. Both lower the chance of winning. These results were steady when I tested on games that the model had not seen before. The model separated winners from losers well, and its predicted chances were reliable. This gives confidence that the patterns here are real and useful.

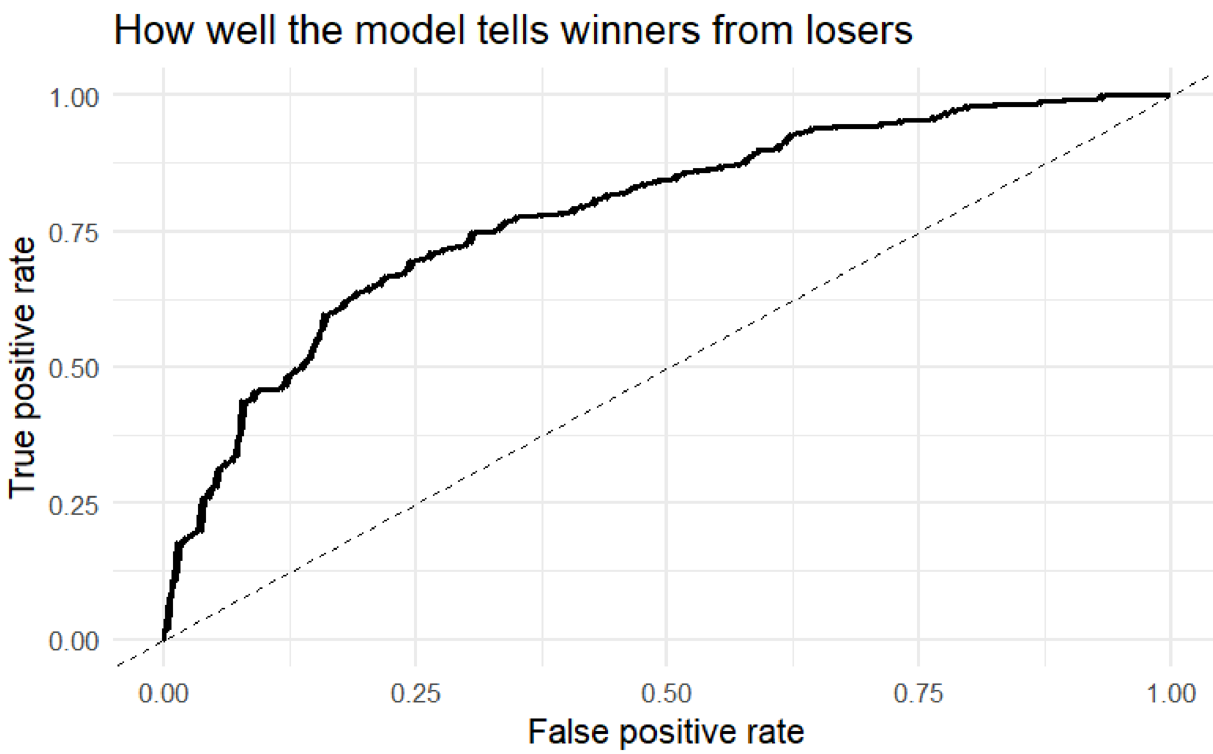

Figure 1: Shows how well the model separates wins from losses across all cutoffs.

This curve shows how well the model can separate games that were wins from games that were losses across all possible decision thresholds. The horizontal axis is the false positive rate, which is how often the model would call a loss a win if you set a certain cutoff. The vertical axis is the true positive rate, which is how often the model correctly calls a win a win at that same cutoff. The dashed diagonal is what random guessing would look like. The black curve bends well above that dashed line, which means the model does a good job telling winners from losers at many different cutoffs. The closer the curve gets to the top left corner, the better the model is.

Understanding the Effects

It is helpful to see each stat and how it relates to the win chance. If we hold the other stats near normal levels and move one stat up and down across realistic values, we can see the model's predicted win chance change. The lines below show that more defensive rebounds and more steals raise the curve, which means higher chances to win. More turnovers or more fouls lower the curve, which means lower chances to win. Blocks do help, but the slope is smaller. This view turns the model into a picture that matches how the game feels.

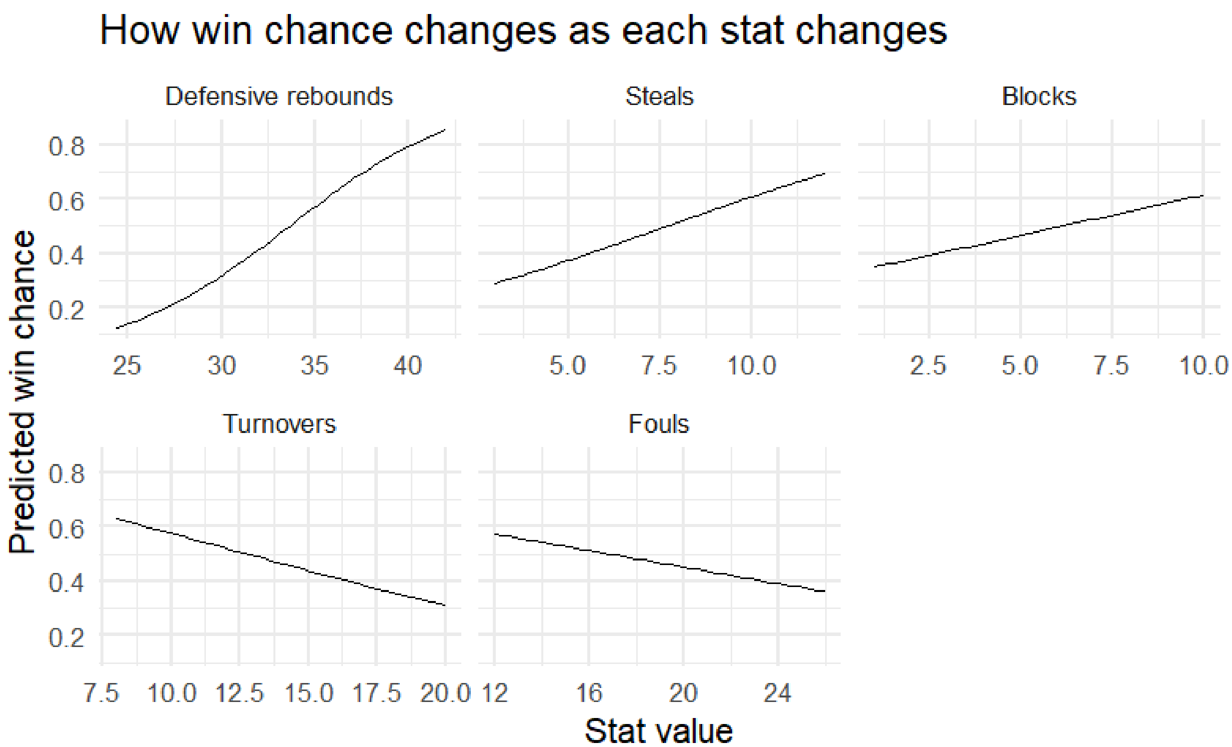

Figure 2: Shows how the predicted win chance changes as each defensive stat moves from low to high.

This chart shows how one defensive stat at a time changes a team's chance to win while the other stats stay at normal levels. Each panel is one stat, and the line shows the model's predicted win chance as that stat moves from a low value to a high value. Lines that climb upward mean "more of this helps," and lines that slope downward mean "more of this hurts." You will see that the lines for defensive rebounds and steals rise the most, which means those two tend to move win chances up the most. The lines for turnovers and fouls go the other way, which means they lower win chances as they increase. The line for blocks does go up, but not as steeply as rebounds and steals, so the effect is smaller.

To put numbers behind those lines, the model's odds ratios show the same pattern in a clear way. An odds ratio above 1 means the stat is tied to better chances to win, and an odds ratio below 1 means the stat is tied to worse chances to win. In this study, defensive rebounds, steals, and blocks all sit above 1, while turnovers and fouls sit below 1. In practical terms, each extra defensive rebound in a game is linked with about 1.23 times higher odds of winning, each extra steal is linked with about 1.20 times higher odds, and each extra block is linked with about 1.14 times higher odds, when the other stats stay the same. These are not promises for any single night, but they are steady signals across many games that match what coaches and players feel on the floor.

Are the Chances Reliable?

It is also important to ask if the model's win chances match what happens on the court. A simple way to see this is to group games by predicted chance and compare the average prediction to the actual win rate in each group. If the points sit near the diagonal line, the model is well calibrated. In this study, the curve stays close to the line. That tells us the numbers are honest and can be used to plan.

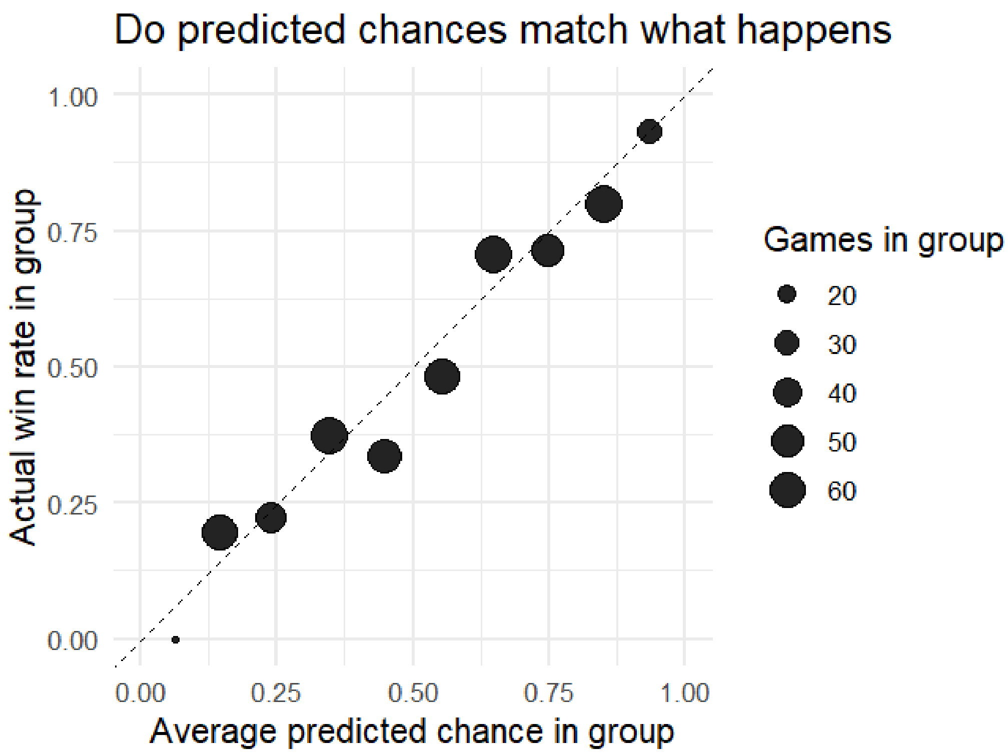

Figure 3: Checks if predicted chances match actual win rates across groups of similar predictions.

This chart is a "reality check" on the model's win chances. Each dot is a group of games with similar predicted chances. The x-axis shows what the model predicted for that group on average, and the y-axis shows how often those teams actually won. The dashed line is perfect honesty: if the model says 70 percent, teams should win about 70 percent of the time. Because the dots sit close to that line across the range, the model's probabilities line up well with what really happened. Bigger dots mean more games in that group, so those points are especially trustworthy.

The results point to a few steady habits. Work to finish defensive possessions. That means strong box outs and team rebounding. Apply controlled pressure to create steals without reaching or fouling. Value the ball and keep outlet options simple to cut turnovers. Defend with your body and keep your hands back to avoid fouls. Do not chase blocks at the cost of position. If the staff tracks these areas each week, the team can see small gains add up. It can also help with rotation choices. For example, if a lineup wins the rebound battle and protects the ball, that is a lineup worth more minutes in close games.

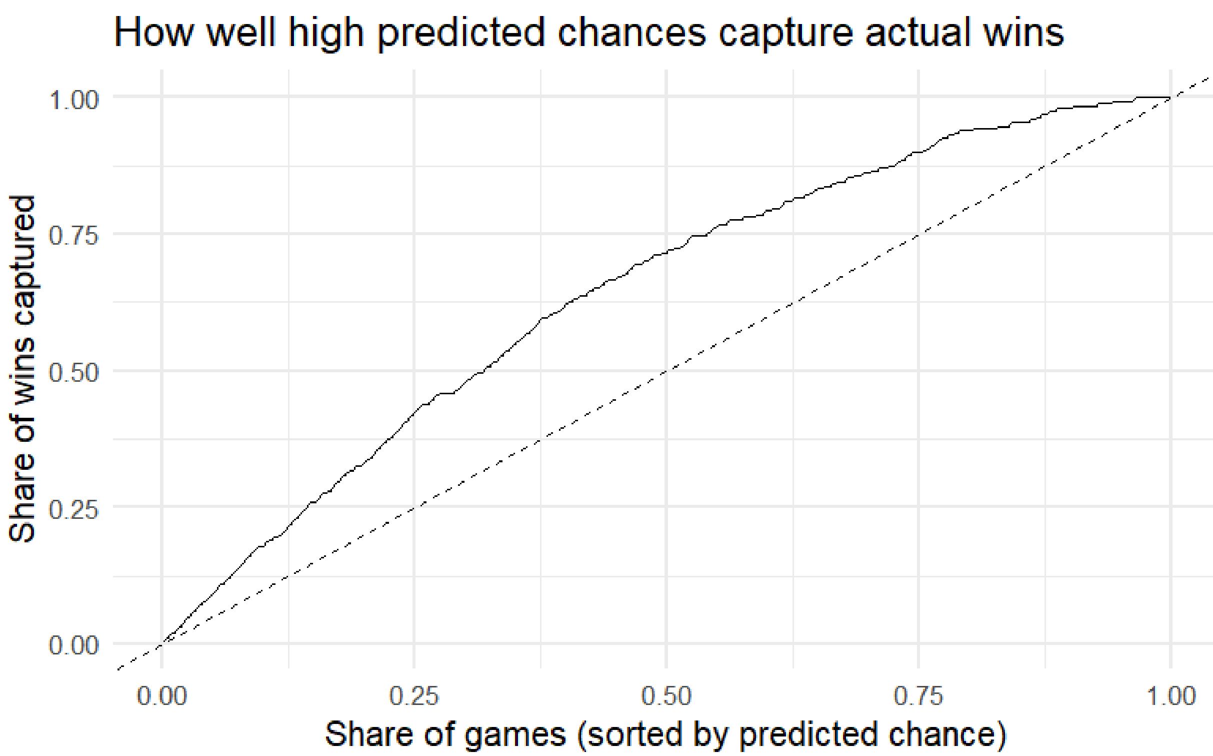

Figure 4: Shows how quickly the top predicted games capture the season's real wins when ranked by probability.

This chart shows how useful the model is for ranking games by their predicted chance to win. The line starts with the games the model thinks are most likely wins and moves down the list. The x-axis is how many games you have taken from the top of that list. The y-axis is how many of the season's total wins you have already captured within those games. The dashed line is what random picking would look like. Because the solid line stays above the dashed line for most of the plot, the model's top-ranked games contain more real wins than you would get by chance. The more the line bows upward, the better the model is at putting true wins near the top of the list.

There are easy ways to make this work even stronger. Adding context such as opponent strength, pace, rest days, and home or away can sharpen the estimates. Turning raw counts into per-possession rates helps compare teams that play fast and teams that play slow. It would also be helpful to compare defense only to offense only and to a combined model. That shows where most of the signal comes from and where the mix is best. A method that respects the fact that each game has two team rows can also improve the error bars. These steps would make the results even more precise while keeping the message simple.

Closing Thoughts

These findings fit how NBA games are won. Teams raise their chances by ending possessions with defensive rebounds and by creating extra possessions with steals, while avoiding turnovers and needless fouls. Blocks help, but the day-to-day drivers are controlling the glass, taking the ball cleanly, and valuing every trip. That is why the model's patterns line up with what coaches emphasize in film and practice.

This also connects to the core ideas of basketball. Winning often comes from small, repeatable habits that change shot volume and shot quality. When you finish stops with a rebound, you prevent second-chance points and often start better offense. When you protect the ball and foul less, you avoid free points and empty trips. Over four quarters, these steady edges add up and help teams manage pace, momentum, and late-game execution.

Results from NBA data will not map perfectly to every high school or college setting, because pace, spacing, shooting, and foul environments differ. Even so, the direction of the lessons travels well. Emphasize finishing the stop, apply smart pressure for steals, and limit giveaways and fouls. Coaches at lower levels can adapt the targets to their roster and competition, but the same habits are likely to lift win chances.