# Luck vs Skill in College Basketball

Jake Browning, Rohan Mawalkar, Seth Peacock

November 2024

# 1 Letter to the Editor

In addition to this report, write a one-page letter to the newspaper chief editor, explaining the main results of the report and suggesting findings that can be communicated with the basketball fans reading the newspaper.

## 2 Overview

Describe the problem, your model, results and how your model performed.

## 3 Introduction

People love sports because they are a showcase of hard work and talent. People also love sports because they are unpredictable. How do we quantify the effect of one versus the other in a basketball game? On the one hand, we often see this framed as a strict dichotomy between luck and skill: sports outcomes depend in some part on one and in some part on the other. However, careful examination reveals that it is rather unclear where to draw the line between luck and skill. For example, if a team happens to study exactly the right film the night before a big game and is able to perfectly foil the other team's plans, is that luck or is it skill? On the one hand, it certaintly takes skill to study your opponent and use this information to your advantage. On the other hand, they could have easily picked the wrong film to study! Instead, we would like to approach this problem using a dichotomy between certainty and uncertainty. This also addresses the issue of a lack of data; there are many aspects of "skill" (such as individual player records) which require more data than we have been able to gather so far.

Our goal here is twofold: first develop a model which predicts the outcome of a basketball game, and second quantify the effect of luck/uncertainty in a basketball game. Since we want to incorporate the effect of uncertainty, our prediction model will contain a stochastic element which we will have to derive. To train our model, we will use data from the regular season games leading up to the 2024 March Madness Tournament, and then test our model's ability to predict the outcome of the tournament. However, we will also discuss additional markers of uncertainty in our data to better understand the effect of uncertainty on a basketball game in a more general sense.

## 4 Methods

Our model will be a modified version of the classic Elo ranking in chess, which we will use along with data from throughout the regular season to assign each team a ranking. (From here on, when we say "Elo ranking" we will mean our modified ranking, and will say "classic Elo ranking" if we mean the original.) Given a pair of teams, these rankings will be used to give us probabilities of an upset (that is, the probability that the team with the lower Elo rating wins).

### 4.1 Our "Elo" Ranking

Traditionally (see [4]), the Elo ranking system first calculates an expected score for the game as a function of the difference in the two teams' Elo ratings. Here, $E$ is the "main" team while $E_{opp}$ is the opposing team:

$$E = \frac{1}{1 + 10^{\gamma_{opp} - \gamma}/c}$$

where $\gamma_n$ is the main team's pre-game Elo score (after game $n$), $\gamma_{opp}$ is the opposing team's pre-game Elo score, and $c, K > 0$ are constants. The main team's Elo rating is updated after their $n+1^{\text{th}}$ game based on the following formula:

$$\gamma_{n+1} = \gamma_n + K * (O - E)$$

Where $O$ is 1 if the team wins and 0 if the team loses. This is essentially performing a stochastic gradient descent (taking a step after each game) for a logistic regression model, where the Elo ratings are the weights of the model [3].

However, this score update does not take into account the score differential of the outcome of the game. Thus we propose a modified Elo system which uses the score outcomes to update a team's Elo score as follows:

$$\gamma_{n+1} = \gamma_n + K^{\frac{S_{\text{win}}}{S_{\text{lose}}}}(O - E)$$

where $S_{\text{win}}$, $S_{\text{lose}}$ are the scores of the winner/loser and the other variables are defined as above. We will calculate $E$ in the same way setting $c$ to the canonical 400 for simplicity, and choose $K$ based on considerations described in the next section near a canonical value of $K \approx 32$ [4].

We considered several other

### 4.2 Bootstrapping

We would now like to verify the accuracy of our Elo system. To do this, we "bootstrap" [2] samples from the regular season games to compare our model's predictions of the probability of an upset to the actual probability of an upset for a given difference in Elo rankings. If our Elo ratings are accurate, we should see these are close to one another.

Note that the probability of an upset obtained via bootstrapping $\hat{p}_U(\Delta_{\text{Elo}})$ could be anything between 0 and 1, while the expected outcome probability from Elo rankings is necessarily between 0 and 1/2. However, we expect that $\hat{p}_U(\Delta_{\text{Elo}})$ should be roughly less than 0 to 1/2, for the same reasons that we hope our Elo expected outcomes $E_U(\Delta_{\text{Elo}})$ are similar to $\hat{p}_U(\Delta_{\text{Elo}})$ for each value of $(\Delta_{\text{Elo}})$.

To quantify the accuracy of our Elo system, we plot $E_U(\Delta_{\text{Elo}})$ against $\hat{p}_U(\Delta_{\text{Elo}})$, fit a linear regression, and calculate the MSE.

### 4.3 Quantifying Uncertainty

To first order, we now also have two (hopefully similar) ways of quantifying the measure of uncertainty in a basketball game: the expected probability of an upset based on Elo rankings (as a function of the difference in Elo rankings), and the bootstrapped estimate for the probability of an upset (as a function of the difference in Elo rankings). However, there is an additional higher order kind of uncertainty, which is the variance in the bootstrapped sample of games with a given score difference.

### 4.4 Weaknesses

### 4.5 Rejected Ideas

- We considered defining the score differential as the absolute value of the difference of the scores. However, this would weight a 20–10 loss as the same as a 90–80 loss, which seems much less accurate to how basketball fans see the game.

## 5 Results

What does your model say about the question you have been given? In particular, you may consider whether incorporating the margin of victory in each game can change your model predictions of the contributions of ability and chance to the sports outcomes. You may also assess how accurately your model could predict the outcome of the 2024 March Madness Tournament.

### 5.1 Verifying Elo Rankings

Here we plot $E_U(\Delta_{\text{Elo}})$ against $\hat{p}_U(\Delta_{\text{Elo}})$ along with the $y = x$ line. If our Elo system has perfectly captured the expected upset probability, then these points should be close to the $y = x$ line. To measure the inaccuracy, we calculate the mean squared error.

When updating our "Elo" scores, we had an artbitrary parameter $K$. We repeated this process with a range of $K$ and see which one minimizes the MSE; the plot above shows the result of the minimized MSE. The plot below shows the effect of $K$ on the MSE. It is important to note that picking the $K$ we did

### 5.2 Predicting March Madness 2024

We would like to see how well our model could have predicted the results of the March Madness 2024 Tournament. To this end, we ran a large number of simulations using our Elo predictions with our predicted chance of upset and seen what percentage of times we correctly predicted the outcome of the tournament.

Of course, due to the uncertainty everyone acknowleges exists, this percentage would be low no matter what. After all, no one has ever correctly guessed the result of the tournament [1]. However, we note that our model (on average) performed better than just picking the winner as the one with the better record.

## 6 Next Steps

There are several things we would have liked to check if we had had more time. For one, it would be interesting to explore the effect of overfitting by the choice of $K$. That is, does the value of $K$ which minimized the MSE for the bootstrapping verification also maximize the likelihood of predicting the actual outcome of the tournament? If not, then we should likely look into a different way of picking $K$ (perhaps one which incorporates past data) as this choice seems to have lead to overfitting on the regular season games.

### 6.1 Additional Data

- We would have liked to incorporate the homefield advantage in the expected score calculated from the Elo system. If we had had this data, we would have added some number $A$ from the expected score of the home team and subtracted it from the expected score of the away team. For example, if we found that home teams on average win 53% of games, then we would add/subtract $A = 0.03$. This way, the expected score of two teams with the same Elo score would be $0.5 \pm 0.03$.
- To pick the $K$ that minimizes the MSE for our Elo system compared to the bootstrapped estimate, we could use data from past seasons.

## References

[1] Christopher Brito. Has anyone ever had a perfect bracket for March Madness? The odds and precedents for NCAA predictions — cbsnews.com. `https://www.cbsnews.com/news/perfect-bracket-march-madness/`. [Accessed 16-11-2024].

[2] Trist'n Joseph. What Is Bootstrapping Statistics? — Built In — builtin.com. `https://builtin.com/data-science/bootstrapping-statistics#:~:text=%E2%80%9CThe%20advantages%20of%20bootstrapping%20are,other%20groups%20of%20sampled%20data.%E2%80%9D`. [Accessed 16-11-2024].

[3] Steven Morse. Elo as a statistical learning model — Steven Morse — stmorse.github.io. `https://stmorse.github.io/journal/Elo.html`. [Accessed 16-11-2024].

[4] Stanislav Stankovic. Elo Rating System — stanislav-stankovic.medium.com. `https://stanislav-stankovic.medium.com/elo-rating-system-6196cc59941e`. [Accessed 16-11-2024].