**A Poisson Betting Model with a Kelly Criterion Element for European Soccer**

Other Sports Track

7WA5RO9BD

1. **Abstract**

Sports betting has experienced a rapid rise in popularity as accessibility and commercialization of daily fantasy and live betting has increased. As sports betting is legalized in different countries and states, we are presented with a new opportunity to create statistical models that we can utilize to predict outcomes of different sporting events that can then be used to find inefficiencies in sportsbooks. In this paper, we attempt to create a model for European Soccer and measure its performance against betting markets to understand if this model can be used to generate profits. The paper shows how we assigned teams an attacking and defensive rating based on their production against their league counterparts which we then utilized to create a Poisson distribution that determined the probability of each event (Win, Lose, Draw). To add another dimension to our model, we used an optimization technique known as Kelly Criterion to determine the optimal amount of money that should be bet on each match. This technique generates bet amounts while also creating a value (KCO Value) that acts as an accurate estimator of the risk associated with each match. By exploring the characteristics of this value, we were able to maximize the success of the model. After running the model for the 2018 and 2019 seasons across the five major European soccer leagues, we can safely say that our model was not only successful in predicting outcomes, but also in generating significant profit yields for a user. A profit percentage of 105.93% was yielded using this model, which implies that a user would pretty much double their money using this model. We also evaluate how the model performs in different leagues to understand which league characteristics benefit the model. The highest profit percentage was seen in the 2019 Premier League season with a profit percentage of 151%. The success of the model can not only help users generate significant profits, but it can also expose certain inefficiencies in the market.

1. **Generating Outcome Probabilities**

The first step when creating our predictive model was to generate probability values for every outcome of the game i.e. home team wining, away team winning, and a draw. To do so we used a Poisson Distribution to generate outcome probabilities. A Poisson Distribution requires a value. This value represents the mean number of events during a specified time interval. We are having the represent the mean number of goals scored by a team in 90 minutes, which is the length of a soccer game.

* 1. **Creating Self-Updating Values for Every Game**

We wanted to ensure our value is adjusted for each team based on their opponent and if a team were to be home or away. This would allow us to create a more accurate prediction of how much a team will score in a particular game.

To do so, our first move was to create an attacking and defending strength for each team. Below are the formulas we use to calculate these strengths:

These strengths were calculated as the season progressed to prevent data leakage. We ensured that we only used data that would be available before the start of the game. These strengths were also calculated with home and away splits, so every team had 4 metrics that updates as the season progressed:

1. Home Attacking Strength
2. Home Defending Strength
3. Away Attacking Strength
4. Away Defending Strength

We use these strengths to calculate each team’s predicted goals for a game. The formula below is used:

This process allows us to create a value that factors in 3 major characteristics:

1. The value accounts for the opponent
2. The value accounts for whether the team is playing home or away
3. The value updates as the season progresses, becoming more accurate

To provide more clarity, we will use the Crystal Palace (H) vs West Brom (A) game as an example during the last game week of the 2018 Premier League season. Below is a calculation of both team’s attacking and defending strengths along with our goal prediction for them:

For this game, our Crystal Palace value is 1.3716 and our West Brom value is 0.7255.

* 1. **. Predicting Probabilities for every Score Line using a Poisson Distribution**

Now that we have our values, we can use the Poisson Distribution to calculate our probabilities.

|  |  |  |
| --- | --- | --- |
|  |  | (1) |

represents the number of events we want to predict. For example, if we would want to predict the probability of a team scoring 3 goals, would take the value of 3. Essentially, to calculate each outcome’s total probability we used the above formula to calculate every possible score line from 0 to 0 to 5 to 5. Using simple probability rules, the probability of each score line can be calculated in the following manner:

We can graphically show both team’s probability to score between 0 to 5 goals. We can also create a result matrix depicting the probability of each score line. This can be seen below:

A close up of a map

Description automatically generated

The total outcome probability is the sum of all the score lines that correlate with that outcome. Below is a formula that depicts this:

For the example mentioned earlier, below are the probabilities we predicted using the model:

To ensure that the model had a baseline to create accurate values for a team’s attacking and defending strength, we used the model for games after game week 13. This resulted in a total sample size of 2368 games across the 10 seasons we ran.

1. **Creating a Bankroll Management System**

After developing the predictive model, the next step was deciding how the bets would be placed and distributed based on our outputs. We believe that rather than simply betting on the most likely outcome, we should have our bet amounts reflect our confidence in the outcome. To do so, we use a Kelly Criterion.

The Kelly Criterion model is a form of probability theory most commonly used by investors that has made its way into the betting world. The goal of the model is to maximize profit while accounting for the risk associated with a lost bet. This is done by maximizing the logarithm of the potential ending bankrolls after the bet is placed. Many online sources choose to simplify the math behind the Kelly Criterion Theory into a generic formula that looks like this:

While this formula does involve some principles of the Kelly Criterion Theory, it does not reflect the theory itself. Also, a soccer match has three outcomes, slightly complicating the math of this simplified two outcome formula. Our goal is to simply maximize the logarithm of our potential bankroll according to the Kelly Criterion Theory.

A screenshot of a cell phone

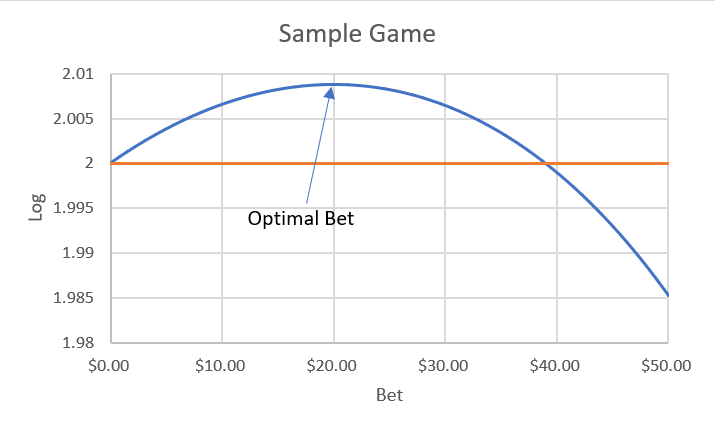
Description automatically generated

Shown above is our optimization for the Crystal Palace vs. West Brom example discussed earlier. The inputs are the odds and probabilities of each possible result as well as the starting bankroll. Using this information, we can calculate our ending bankroll for each possible event and then take the logarithm of that. Lastly, the “objective” is a weighted average of the logarithms and their associated probabilities. Once this information is set, we use Microsoft Excel’s solver add-in to maximize the objective cell by changing the “Bet Amount” cells. A typical output can be seen below:

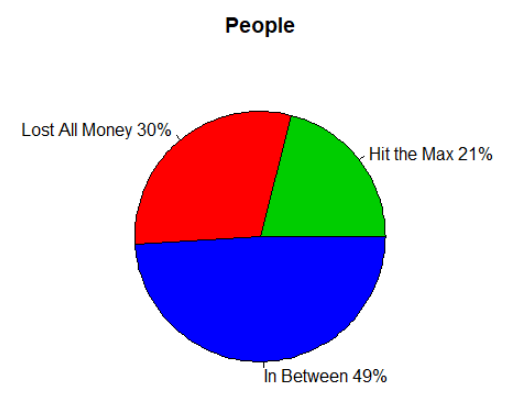
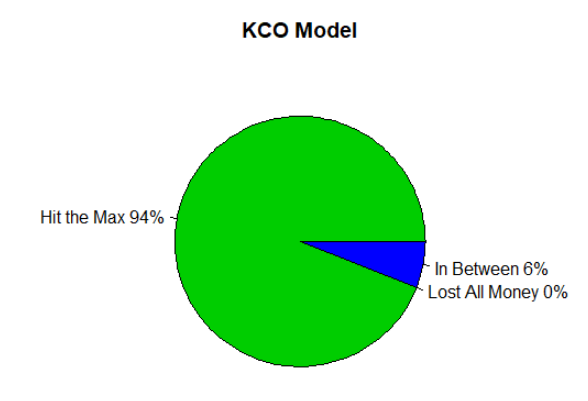
A screenshot of a cell phone

Description automatically generated

Based on these results, the model suggests a bet of $22.34 on Crystal Palace to win and a bet of $1.65 on West Brom to win are the optimal bets to place that maximize long term profitability. The result of the game was 2 – 0 in favor of Crystal Palace, hence we would have profited $16.22 for this bet.



The above graph shows how the objective cell interacts with different bet amounts for a sample game (blue) in which the team being bet on has +100 betting odds and a 60% chance to win the game. It follows somewhat of a quadratic form that varies in shape depending on the betting odds and probability of winning. The left side of the curve indicates bet amounts that are profitable, but not as profitable as higher bet amounts. The right side of the curve indicates higher bet amounts that would be considered higher risk bets that would prohibit long term success. The orange line indicates the objective number if no bet is placed. In other words, bets in which the blue line is below the orange line should not be placed. As mentioned before, Solver essentially looks to find the maximum of the blue line.

This method of risk management and profit maximization has been proven very effective in a study conducted by Victor Haghani and Richard Dewey. In the study they gathered 61 “young, quantitatively trained men and women” to play a game. In the game, each person was given $25 and 30 minutes (about 300 tosses) to place bets on a coin that would land on heads 60% of the time. Prizes were capped at $250. The results were interesting, even somewhat comical. Only 13 people (21%) hit the maximum and 18 (30%) lost all of their money in the end. Furthermore, around 67% of the test group bet on tails at some point, despite it only having a 40% chance to win. When the Kelly Criterion method is used, a user would hit the $250 cap 94% of the time, showing the strength of this methodology relative to the average bettor.

Before moving on to our results, it is important to point out two more aspects of this methodology. The first is how to interpret the “objective” number produced by the optimization. As mentioned before, the objective number is the product of the logarithm of all possible ending bankrolls. Since we assumed a starting bankroll of $100, the objective cell when placing no bet is 2 since there is a 100% chance of ending with $100 and log(2) = 100. After Solver finds the optimal bet, the objective cell becomes a representation of our confidence in profitability with a value of 2.1 having a lot of confidence relative to a value barely above 2. Lastly, there are some games in which a bet might be placed on a team to win as well as a smaller bet on a draw. This occurs as a form of hedging that reduces the potential loss of our bet not hitting.

1. **Model Calibration**

This is the first paragraph of the body of text under the first subsection of the first main section.

**4.1. Brier Scire**

This is the first paragraph of the body of text under the first subsection of the first main section.

1. **Model Evaluation**

In this section, we will observe how successful the model was in terms of overall profit. It

is important to note that the Kelly Criterion suggests different amounts to bet for each game

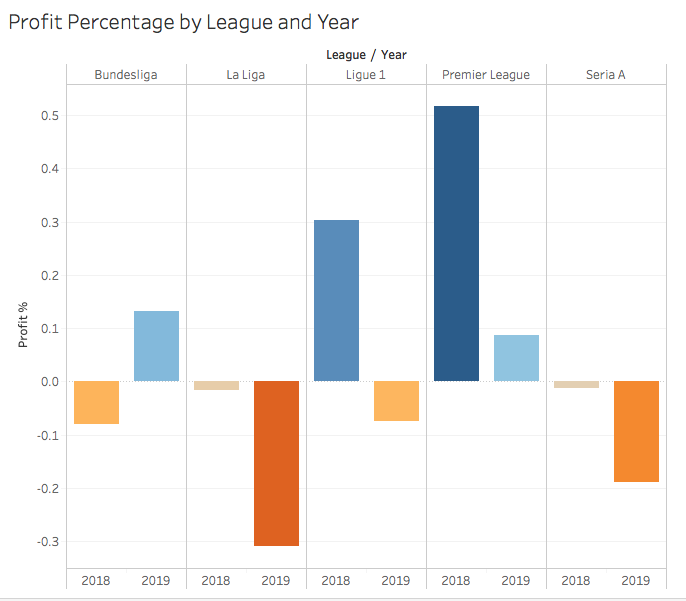
based on the risk, hence we decided to use profit percentage as an indicator for the model’s

success in addition to profit.

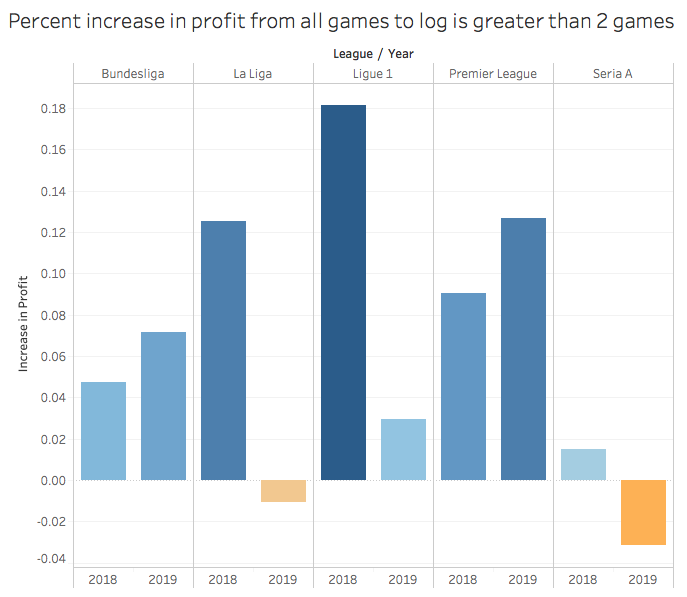
**5.1. Comparing Our Poisson Probabilities with Implied Probabilities**

This is the first paragraph of the body of text under the first subsection of the first main section.

**5.2. Evaluating Risk Using our Generated KCO Values**

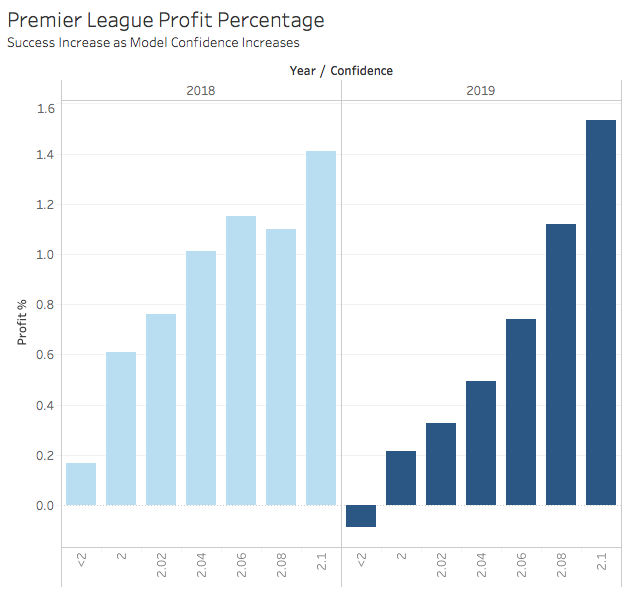


Above depicts our results if we were to follow the model and bet on every game. The Premier League and Ligue 1 in 2018 were highly successful using our model. Specifically, the Premier League in 2018 generated $4,966.72 at 52% profit. Ligue 1 in 2018 had a total profit of $2,635.81 at 30 %. While this overall results of the ten samples seasons were unsatisfactory, we were able to discover that the success rate and profit can be significantly increased when filtering out games that had a KCO generated log value lower than 2.



When only looking at games greater than 2, eight out of the ten leagues experience an increase in profit percentage. Ligue 1 in 2019, Bundesliga in 2018, and Serie A in 2018 all went from negative or breakeven to positive dollar profit when only betting on confident games. In other words, the amount of risk associated with higher logarithm values is less than when a game is at or below 2. In various instance the model usually hedges its bet by throwing a small amount on the draw, allowing for either a lower profit or a smaller loss if a draw occurs rather than the predicted team winning.

To further depict the model’s success and the KCO value’s accuracy as a risk estimator, we can investigate how the success increases as the model’s confidence increases. Below shows increments of 0.2 for our KCO value and how that impacts the profit percentage for the Premier League.



As the confidence got closer to 2.1 in 2018 and 2019, the profit percentage gradually increased all the way up to 141% and 153% respectively when only betting on the games our model was most confident in.

**5.3. Final Observed Profit with our Model**

This is the first paragraph of the body of text under the first subsection of the first main section.

1. **Possible Improvements**

**5.1. Bivariate and Zero Inflated Poisson Distributions**

This is the first paragraph of the body of text under the first subsection of the first main section.

**5.2. Factoring Lineup Selections within Our Attacking and Defending Strengths**

This is the first paragraph of the body of text under the first subsection of the first main section.

1. **Conclusion**

When appropriately using past data to predict game outcome probabilities and then allocating funds efficiently to minimize risk it is clear that there is money to be made betting in European soccer. SUtilizing our predictive model and the Kelly Criterion model, we found success in many leagues such as the Premier League, Ligue 1, and Bundesliga. However, there was still a sizable amount of loss being incurred. This led to the filtering of our KCO log values in each respective league. We found the most success and greatest profit for all games where the KCO log value was close to 2.1 with a profit percentage increasing up to 153% in the Premier League. As discussed above, there are always improvements that can be made with our model, including different methods following each league’s dynamics. We hope to expand our model and continue to grow those profit percentages, finding the most accurate and profitable way to predict European soccer match outcomes.

**References**

[1] Reference #1 cited using any mainstream citation style (e.g. APA, MLA).

[2] Reference #2

.

.

.

[n] Reference #n

**Appendix**

An appendix is not required, but if you have one please include it here.