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

Kushal Shah kshah07@syr.edu, James Hyman jmhyman@syr.edu, Dominic Samangy, dsamangy@syr.edu
Syracuse University Soccer Analytics Club

Research Question

How can Data Modeling be combined with Risk Optimization to Maximize Profits from European Football Betting?

Data Overview and Poisson Distributions

We utilized 2 years of data from Europe's top 5 leagues providing us with 10 seasons worth of data. This included match statistics (Goals, Free Kicks, etc.) and odds for every game.

We generated a λ value for every game that represents the number of goals we predicted each team to score in that game. We created 4 models to generate λ values:

1. Goals Model: Using Goals For & Goals Against we created Attacking/Defending Strengths to generate λ values
2. xG Model: Using xG For & xG Against we created Attacking/Defending Strengths to generate λ values
3. Linear Regression: A Linear Regression using in-game statistics we calculated λ values
4. Random Forest: A Random Forest using in-game statistics we calculated λ values

With these λ values we generated probability values using a Poisson Distribution:

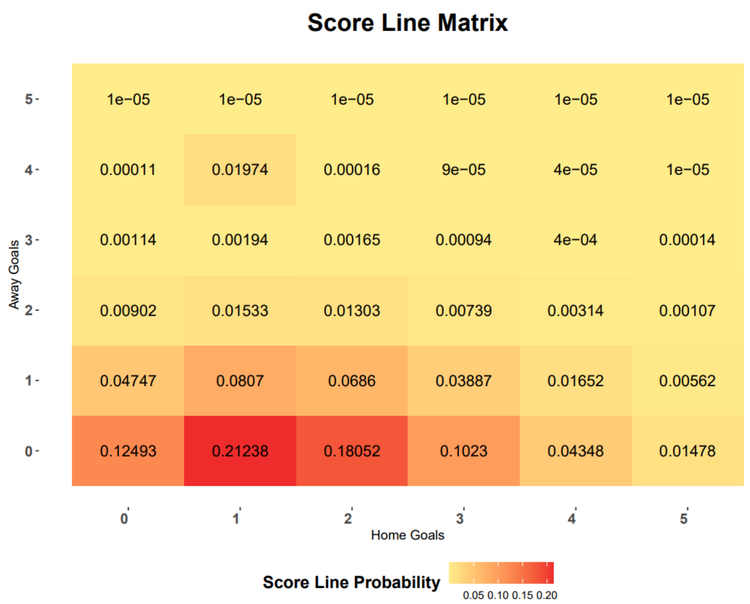
$$P(x) = \frac{e^{-\lambda} \times \lambda^{x}}{x!}$$

We calculated the probability of each score line from 0-0 to 5-5 using basic probability.

Using λ values to create a Score Line Matrix

Here is a Score Line Matrix for a match between West Brom & Crystal Palace

The rows correspond to the number of goals scored by West Brom (Away) & the columns correspond to the number of goals scored by Crystal Palace (Home). We can see that the model believes the most likely score line is 1 – 0 which has a 21.238% (0.31056 × 0.68386) chance of occurring.



The probability of each event was the summation of the score lines corresponding to the respective event. Here is the formula we used to calculate Home Win Probability for a match:

$$HomeWinProb = \sum Probabilities\ of\ Scorelines\ with\ Home\ Team\ Winning$$

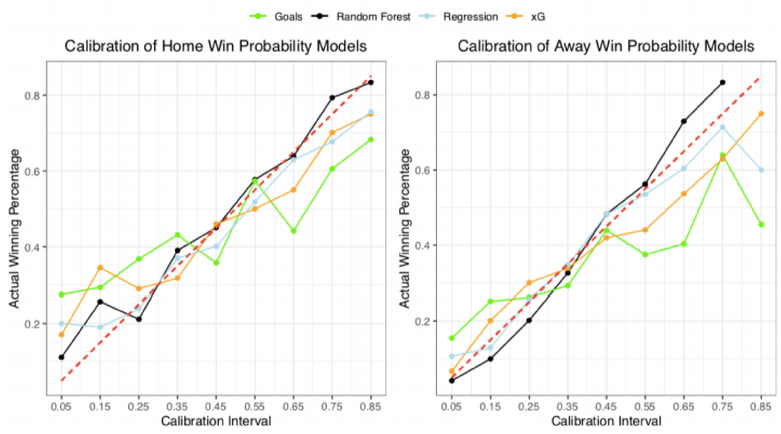
Model Selection and Risk Optimization with the Kelly Criterion

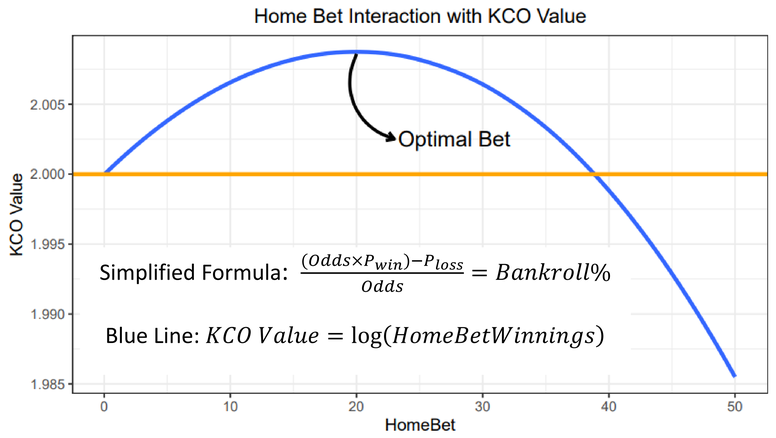
We calibrated the model using a Brier Score to determine which model generated the most accurate λ value.

$$Brier\ Score = \frac{1}{n}\sum_{i=1}^{n}(p_i - o_i)^2$$

We are using this score to measure if the outcomes we predict with a certain probability are occurring in that proportion. For example, among all the events for the rest of the season that we predicted to occur with a probability of 0.6 & 0.7, we would want to be correct around 65% (0.65) of the time. We analyzed each model from week 1 through 13.

Best Performing Model: Random Forest Model Game Week 13 Onwards.





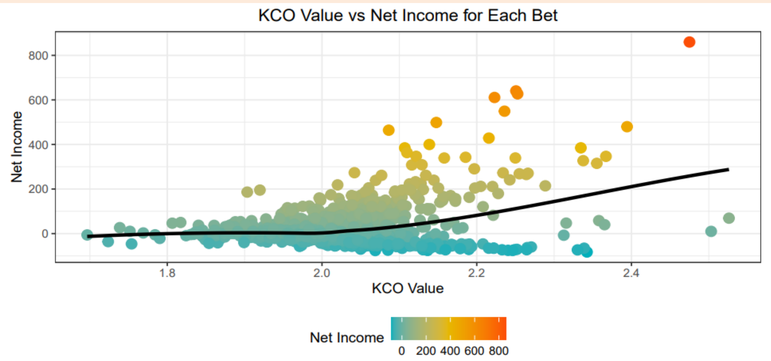
The Kelly Criterion is a probability theory that we adapted to manage our bankroll. We use this theory to place the optimal amount of money that maximizes profits & minimizes risk. The formula above shows the interaction of the theory with a 2-outcome bet.

Since a soccer match has 3 outcomes, the math of this simplified 2 outcome formula is complicated. Our goal is to maximize the log of our potential bankroll according to the Kelly Criterion Theory. Hence, we weight each bet by their likelihood creating an Expected KCO value highlighted below:

$$Expected\ KCO = (HomeWinProb \times log(HW)) + (AwayWinProb \times log(AW)) + (DrawProb \times log(DW))$$

The expected KCO value acts a measure of risk, with bets of a higher KCO value being considered less risky.

Model Results ($$$)



We can clearly see for every season that the overall profit drastically improves as we select bets with higher KCO values. While bets with a KCO value greater than 2.1 have the greatest ROI, they occur at a rarity & hence the maximum volume of profit is achieved from bets with a KCO value greater than 2.

On the right we can see a table with the profits from each league based on their KCO Value. We have highlighted the seasons with the greatest Profit Percentage.

League	KCO Value Greater than 2		KCO Value Greater than 2.1	
	Amount Risked	Amount Profited	Amount Risked	Amount Profited
Bundesliga 2018	$3251.95	$1273.59	$454.73	$616.80
Bundesliga 2019	$4705.58	$7780.69	$1720.42	$5014.22
La Liga 2018	$4633.20	$1363.87	$1098.77	$703.33
La Liga 2019	$5517.66	$2847.05	$1215.85	$2555.30
Ligue 1 2018	$4727.76	$1812.51	$894.89	$1117.37
Ligue 1 2019	$4136.08	$1555.22	$674.37	$519.70
Premier League 2018	$5261.52	$3483.21	$1148.28	$933.22
Premier League 2019	$5464.42	$2060.52	$897.53	$176.17
Serie A 2018	$4519.20	$796.81	$1178.63	$1200.40
Serie A 2019	$4070.81	$804.77	$704.14	$625.49