Assessing the Accuracy of Betting Odds in European Football

Joseph Pym

Student ID: 8404110

Supervisors: Dr. Maha Moustafa & Dr. Alun Owen

Presented 22nd April 2021

Coventry University

§1: INTRODUCTION

Background Information — Football in Europe

Coventry University

Two groups of football leagues were considered in this project

1. *Elite* Leagues

Based on the UEFA Country Coefficients

🇪🇸 92.3	🏴 90.7	🇮🇹 72.3
🇩🇪 71.9	🇫🇷 54.9	🇵🇹 47.4

2. English & Scottish Pyramids

The leagues with > 3 datasets per year on
`football-data.co.uk`

Premier League Championship	1	Premiership
League One League Two	2	Championship
National League	3	League One League Two

Based On

- Average Club Revenue
- Fully-Professional Lower Leagues
- Leagues with Non-Professional Teams

Background Information — Gambling

Coventry University

Probabilities, Odds & Gambling

$$\mathbb{P}(\text{Event}_i) = p_i \in [0,1]$$

$$\text{Underlying Probability} = \frac{1}{O_i}$$

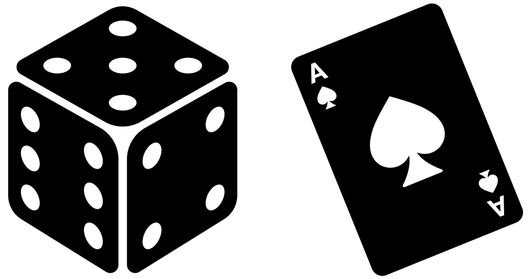
Once found, **normalise** the underlying probability to find the bookmaker **consensus** probability

Markets of Interest

1X2 — Betting on the explicit outcome of the match being a Home Win, Draw, or an Away Win

Under/Over 2.5 Goals — Are the total match goals Under or Over 2.5?

Asian Handicap — A handicap is applied to the home team, aiming for a 50% chance of either side winning.



Coventry University

Aims, Objectives & Methods

What?

- Accuracy of the 1X2 odds offered in *elite* leagues;
  Accuracy over time;
  Competitive Balance
- Different Levels & Different Markets;
  Overround (commission)
- Development of a betting algorithm and comparison against randomly placed bets

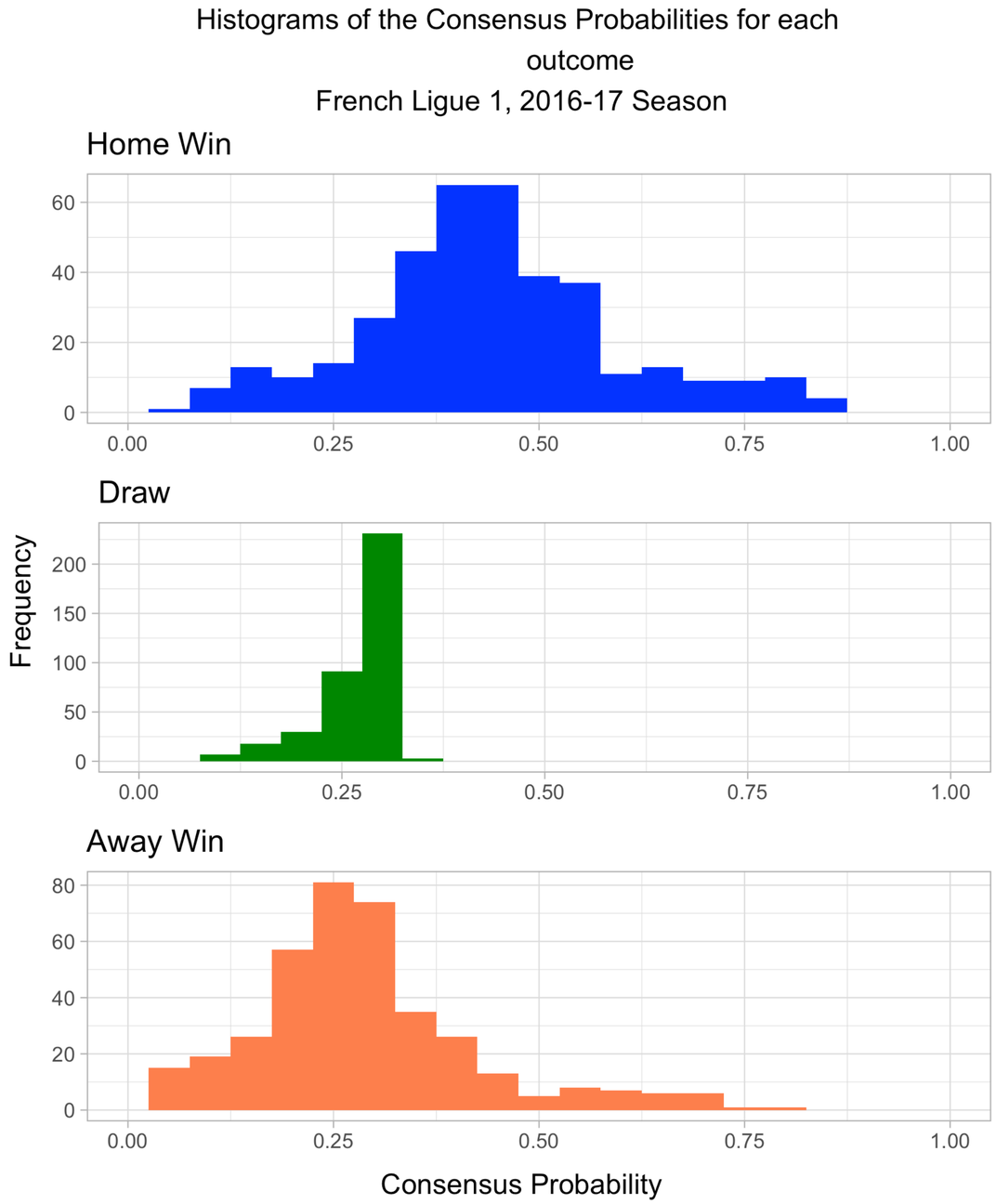
How?

- Initial and Exploratory Analysis; Visual Analysis; Correlation Analysis, ETC.
- Data: `football-data.co.uk`
- Software Language: R

Coventry University

§2: *ELITE* LEAGUES

Assessing the accuracy of betting odds in *elite* European leagues from 2005-2020 in the 1X2 betting market.



Initial DA

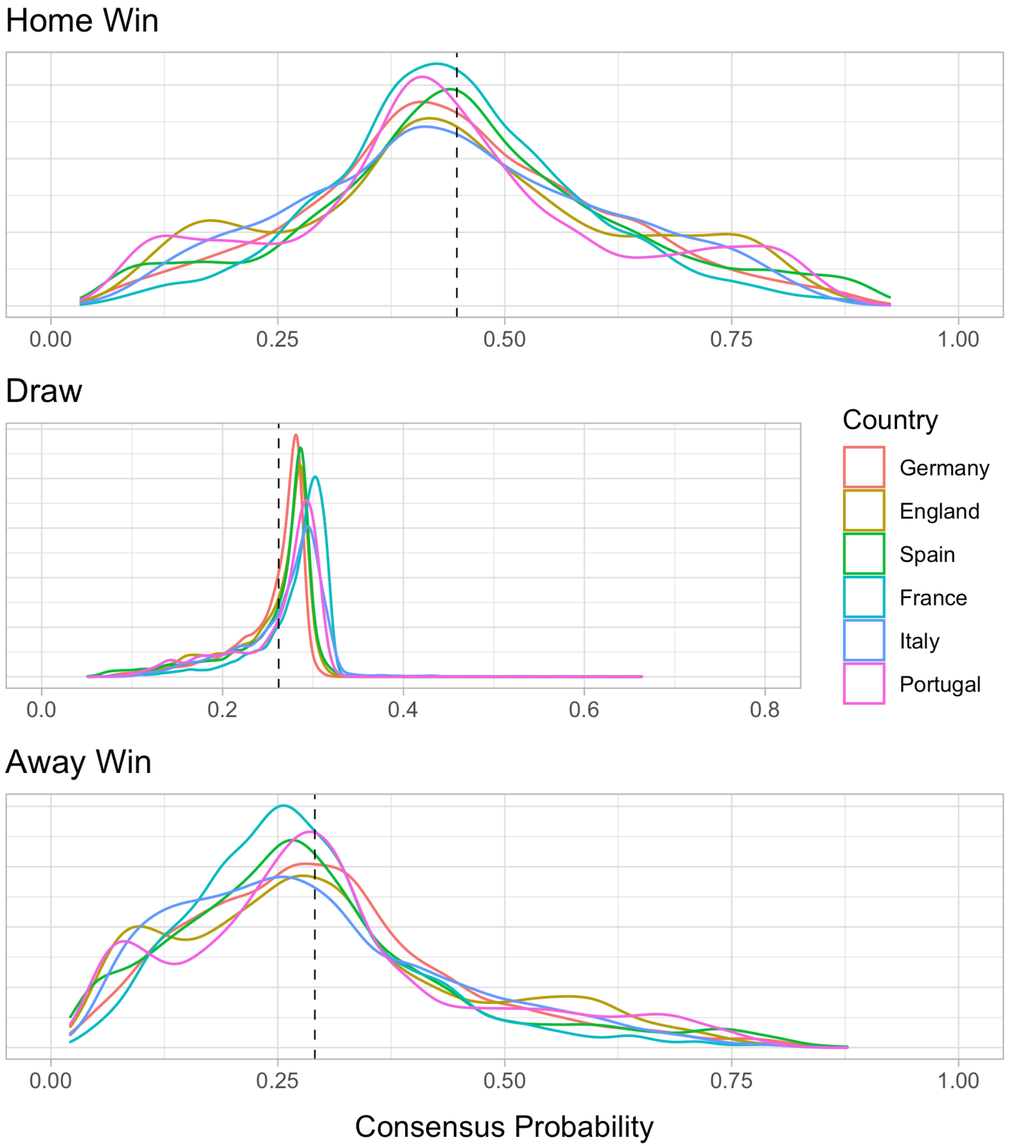
- Ran simple steps on a smaller dataset to ensure the data reads correctly and behaves as expected
- Randomly chose **French Ligue Une, 2016/17**
- ***Observed Probability*** $= \dfrac{\text{Games with that outcome}}{\text{Total number of games}}$

	Home Win	*Draw*	*Away Win*
Mean Consensus Probability μ_i	0.4414	0.2701	0.2886
Observed Probability	0.4895	0.2474	0.2632
Consensus Standard Deviation σ_i	0.1540	0.0469	0.1366

Coventry University

Exploratory DA

Density Plots for the Cons. Probability for each outcome
Elite European Leagues, 2005-2020



	Home Win	*Draw*	*Away Win*
Mean Consensus Probability μ_i	0.4472	0.2620	0.2908
Observed Probability	0.4589	0.2566	0.2845
Consensus Standard Deviation σ_i	0.1714	0.0478	0.1536

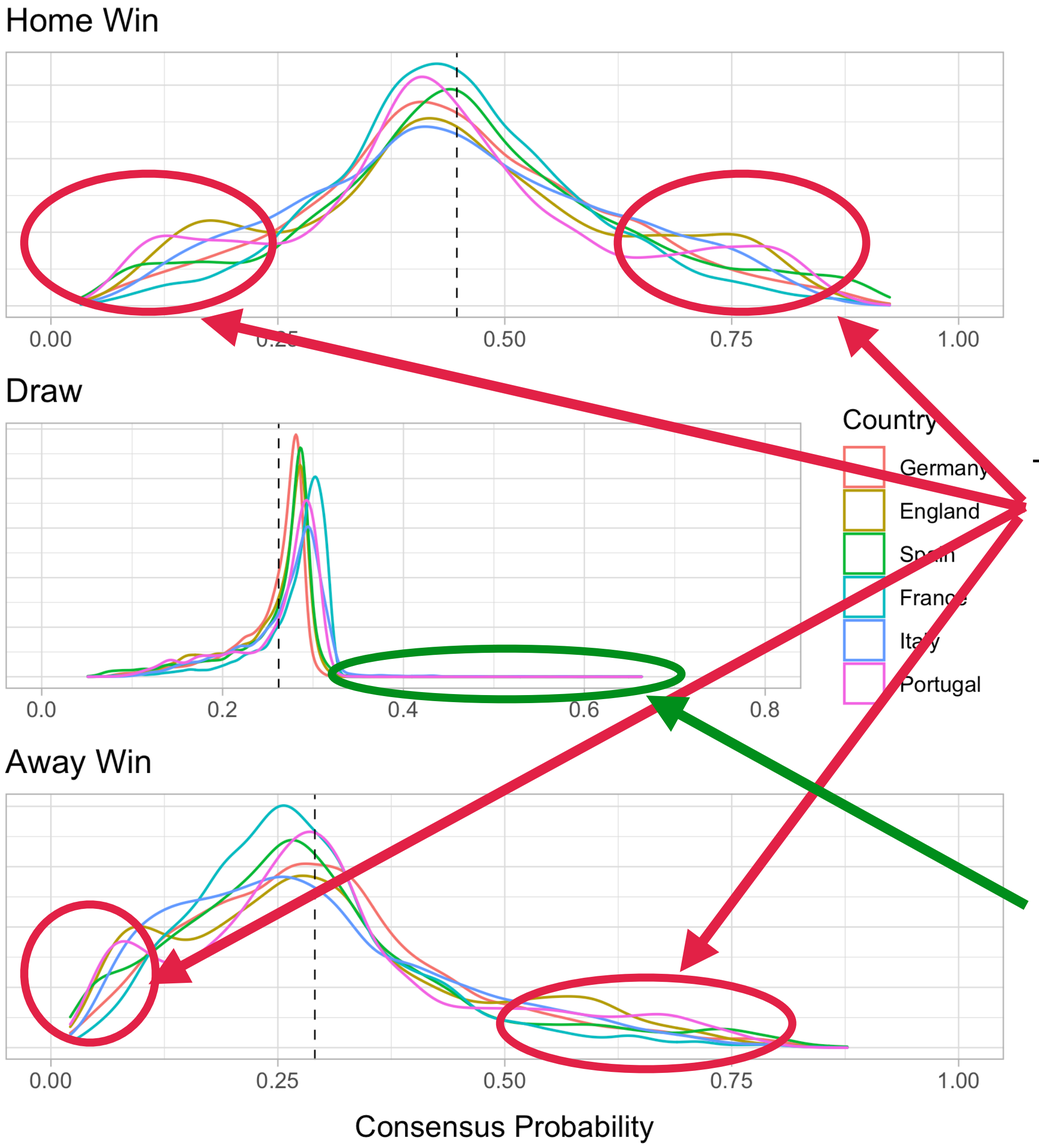
These plots show how the market probabilities are **distributed** and highlights differences across leagues.

Motivated areas for further research.

Coventry University

Exploratory DA

Density Plots for the Cons. Probability for each outcome
Elite European Leagues, 2005-2020



	Home Win	*Draw*	*Away Win*
Mean Consensus Probability μ_i	0.4472	0.2620	0.2908
Observed Probability	0.4589	0.2566	0.2845
Consensus Standard Deviation σ_i	0.1714	0.0478	0.1536

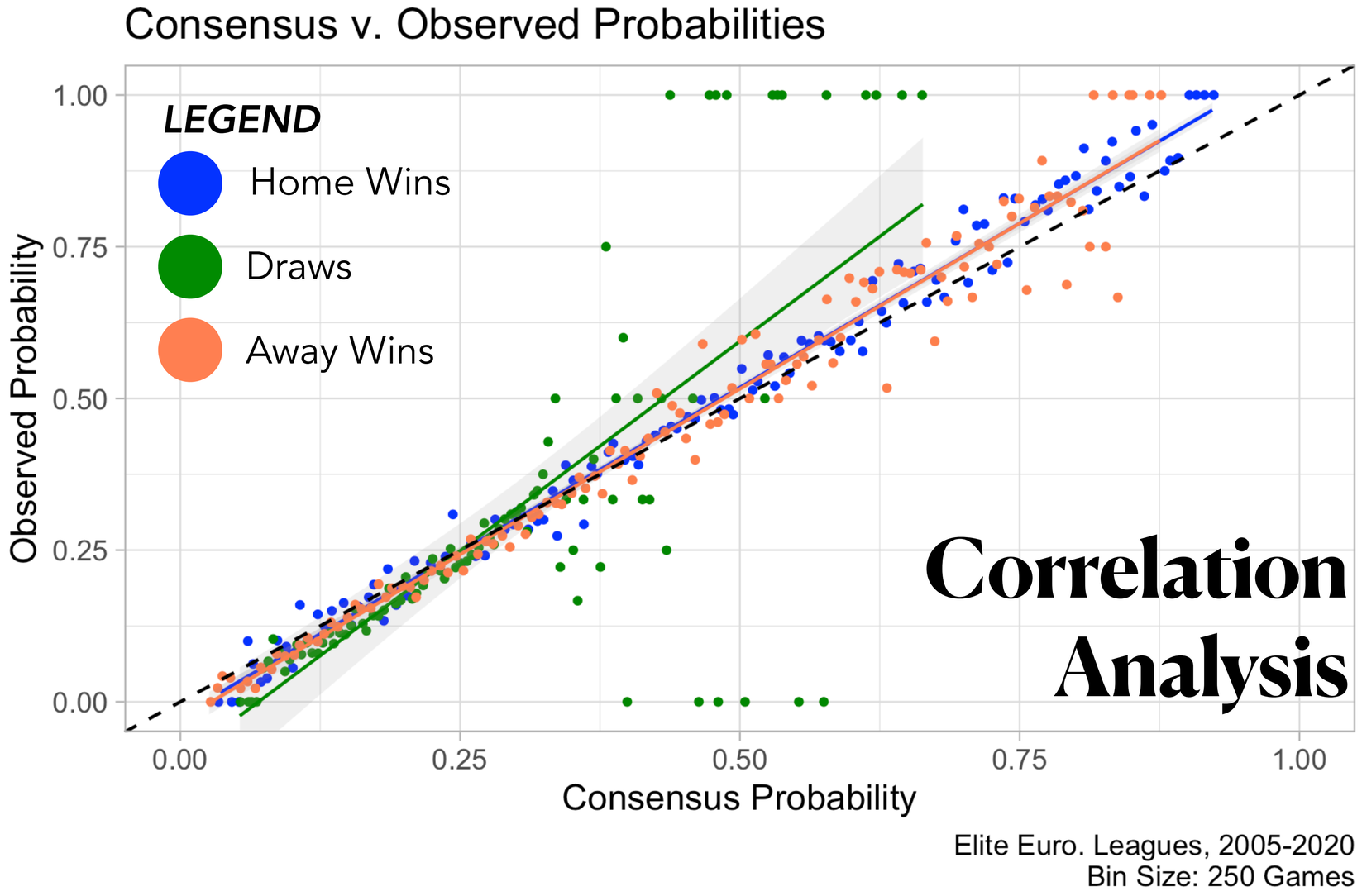
The English, Spanish and Portuguese leagues display a **trimodal** distribution

High P(Draw) => Match fixing or teams playing for mutually acceptable scores *(small number of matches, so we can ignore)*

These plots show how the market probabilities are **distributed** and highlights differences across leagues.

Motivated areas for further research.

Coventry University



Data was *binned* into groups with a similar consensus probability (250 games +) (x axis)

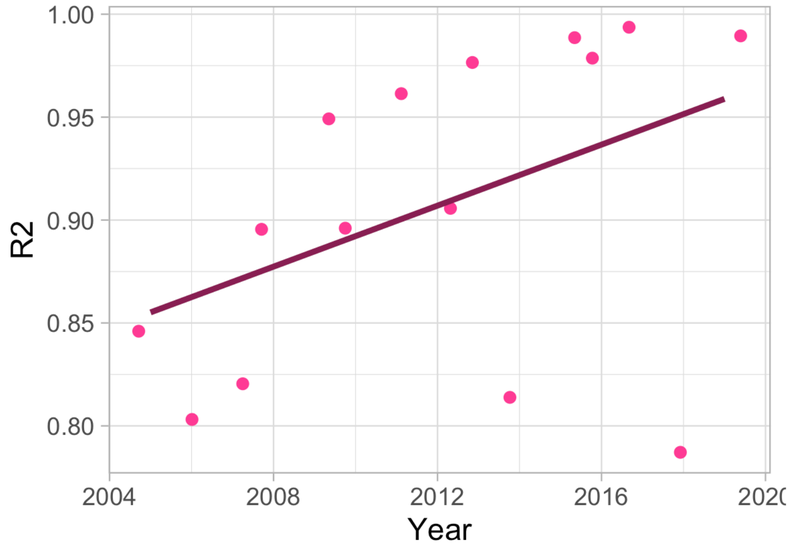
The observed proportion, P(Outcome) for each bin is found, and plotted (*y* axis)

A **linear model** is created to find values of R^2 and RMSE, both for each outcome and overall

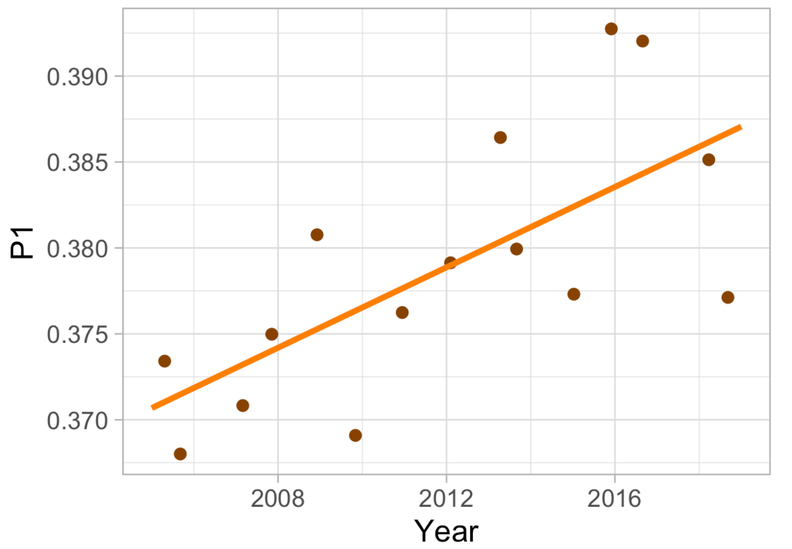
	Home Wins	Draws	Away Wins	Overall
R^2	0.98665	0.52008	0.96411	**0.83832**
RMSE	0.03241	0.20767	0.05166	**0.11768**

Coventry University

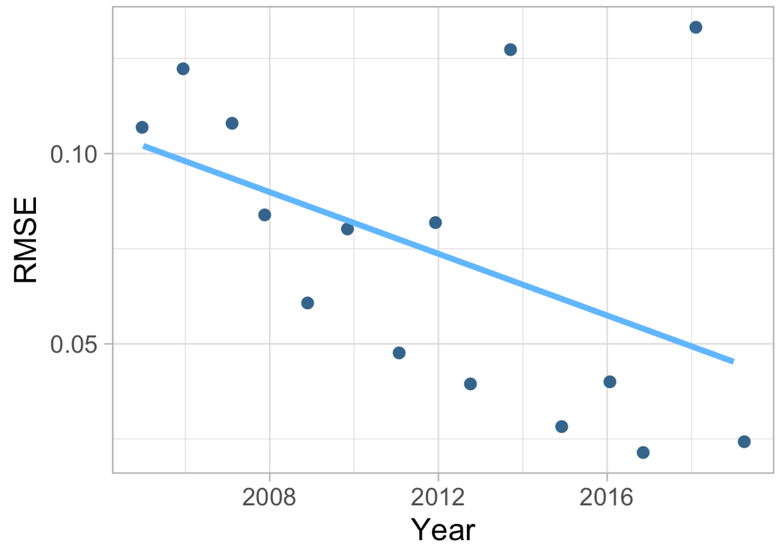
Statistics for **Accuracy over Time**

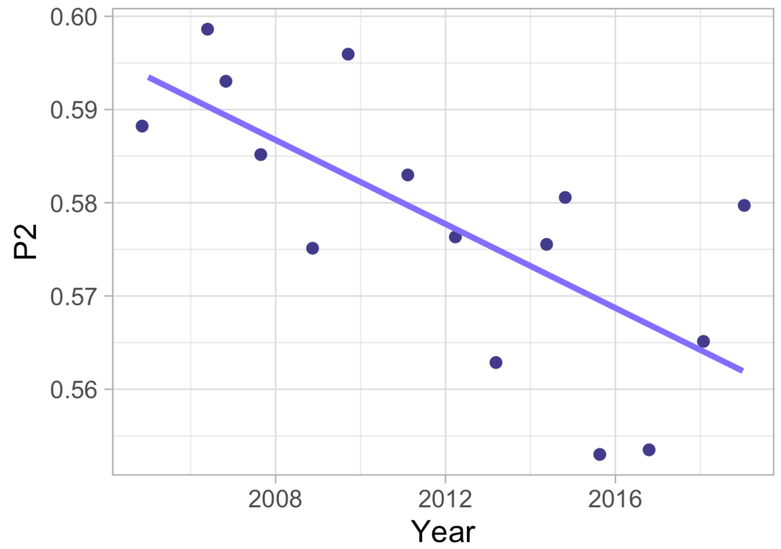


More Accurate



More Accurate





R^2, RMSE

Based on the linear models created in correlation analysis

Differs depending on *bin size*

P_1, P_2

Based on the *correct* and *incorrect* consensus probabilities

Independent of *bin size*

Comparative measure

The effect of Competitive Balance

Coventry University

Ranked least (1st) to most (6th) competitive

Ranked best performance (1st) to worst (6th)

COUNTRY	NAMSI	KAPPA κ	GINI	AVE.	R^2	RMSE	P_1	P_2	AVE.
PORTUGAL	0.505 (1st)	4.07 (1st)	0.898 (1st)	1	0.982 (2nd)	0.036 (2nd)	0.389 (1st)	0.560 (1st)	1.5
ITALY	0.418 (2nd)	5.36 (3rd)	0.737 (5th)	3.33	0.986 (1st)	0.033 (1st)	0.383 (3rd)	0.571 (4th)	2.25
ENGLAND	0.372 (4th)	5.79 (5th)	0.826 (3rd)	4	0.968 (5th)	0.048 (5th)	0.385 (2nd)	0.567 (2nd)	3.5
SPAIN	0.364 (5th)	5.07 (2nd)	0.861 (2nd)	3	0.970 (4th)	0.047 (5th)	0.383 (4th)	0.570 (3rd)	3.75
FRANCE	0.342 (6th)	6.00 (6th)	0.784 (4th)	5.33	0.979 (3rd)	0.038 (3rd)	0.366 (6th)	0.602 (6th)	4.5
GERMANY	0.374 (3rd)	5.71 (4th)	0.721 (6th)	4.33	0.955 (6th)	0.055 (6th)	0.370 (5th)	0.594 (5th)	5.5

$$\text{NAMSI} = \frac{\sqrt{\frac{\sum_{i=1}^{n}(W_i - 0.5)^2}{n}}}{\sqrt{\frac{\sum_{i=1}^{n}(C_i - 0.5)^2}{n}}} = \frac{\sigma\text{Season}}{\sigma\text{Certainty}}$$

Kappa κ is the number of unique teams in the top 3 places across the last 3 years

Gini coefficient (based on Lorenz curve) measures inequality: High scores = More inequality

Portugal is the most competitively **imbalanced** and is where bookmakers perform best

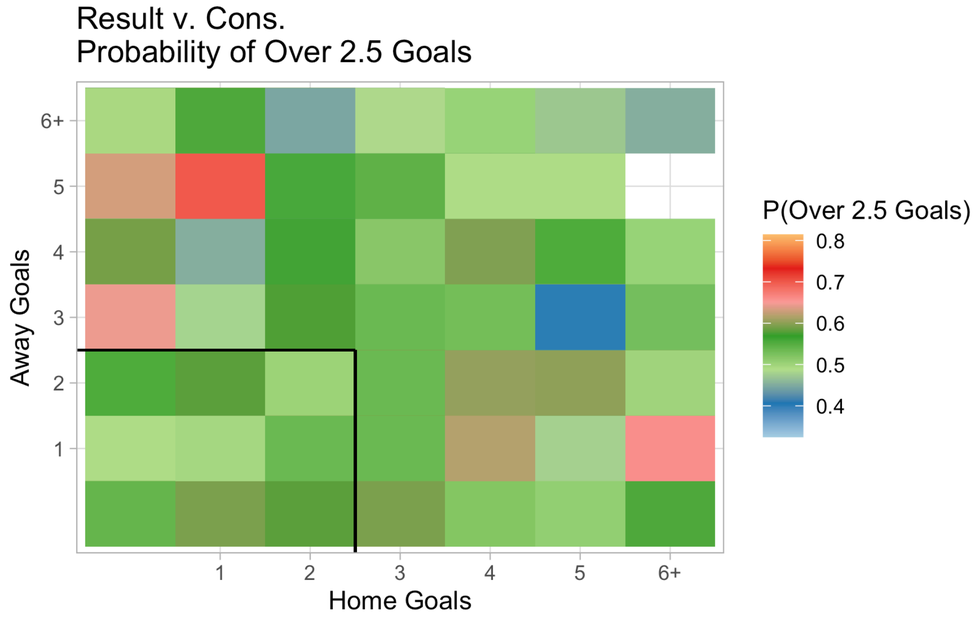
Germany and **France** are the most well-balanced leagues, and are where bookmakers perform worst

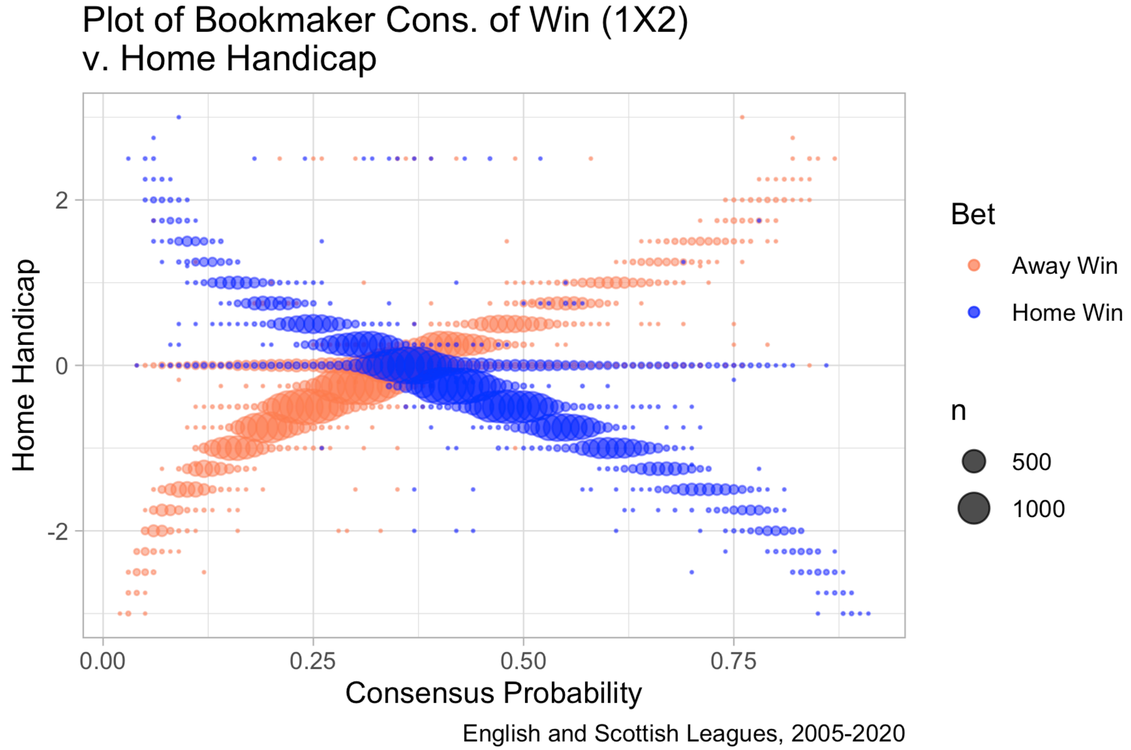
Coventry University

§3: ENGLISH & SCOTTISH LEAGUES

Assessing the accuracy of betting odds in the English & Scottish football pyramids from 2005-2020, across three betting markets.

Coventry University





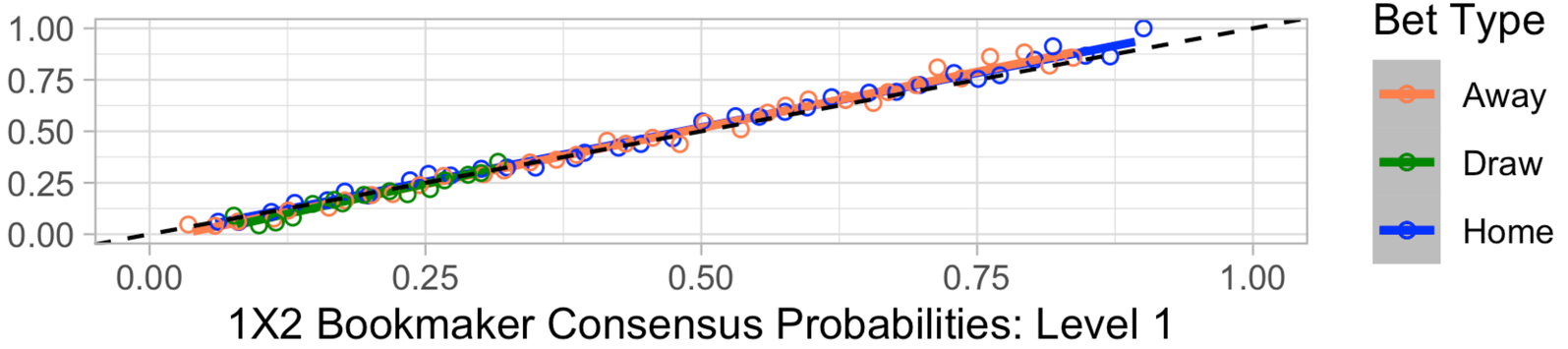
English and Scottish Leagues, 2005-2020

Findings

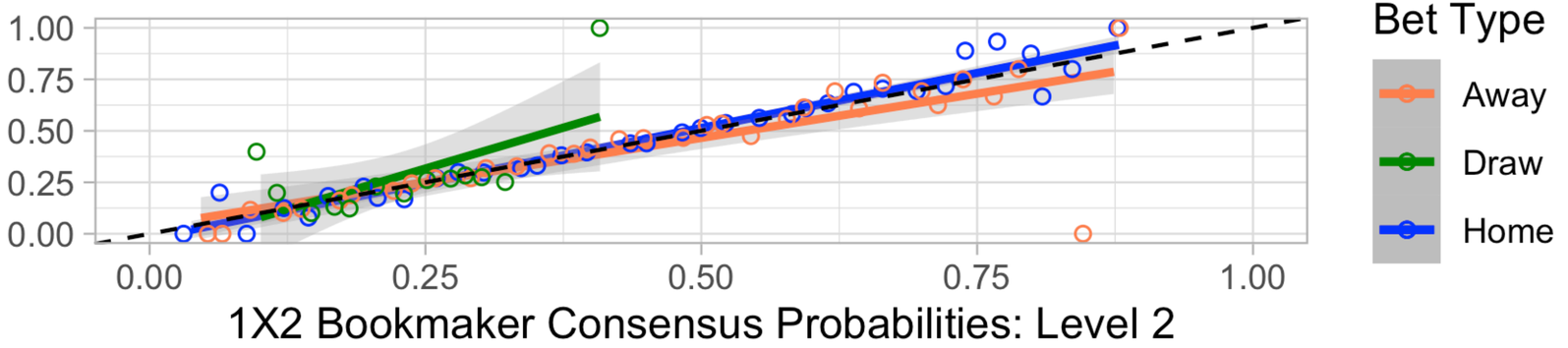
- **Under/Over 2.5 Goals market** is **not accurate** using R^2 and RMSE as measures
- The **Asian Handicap** handicap is **well-placed**, levelling the playing field (probabilities are close to 0.5 for the majority of matches)
- Level 2 matches were **less accurate** and **less varied** than Levels 1 and 3 across in the **1x2 Market**

1X2 Observed vs. Consensus Probabilities

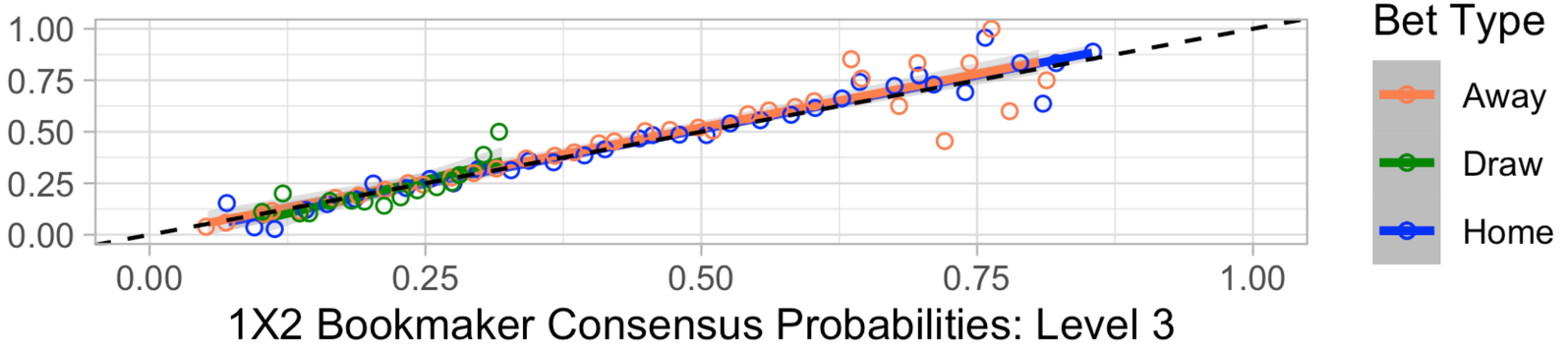
Coventry University



Level 1



Level 2



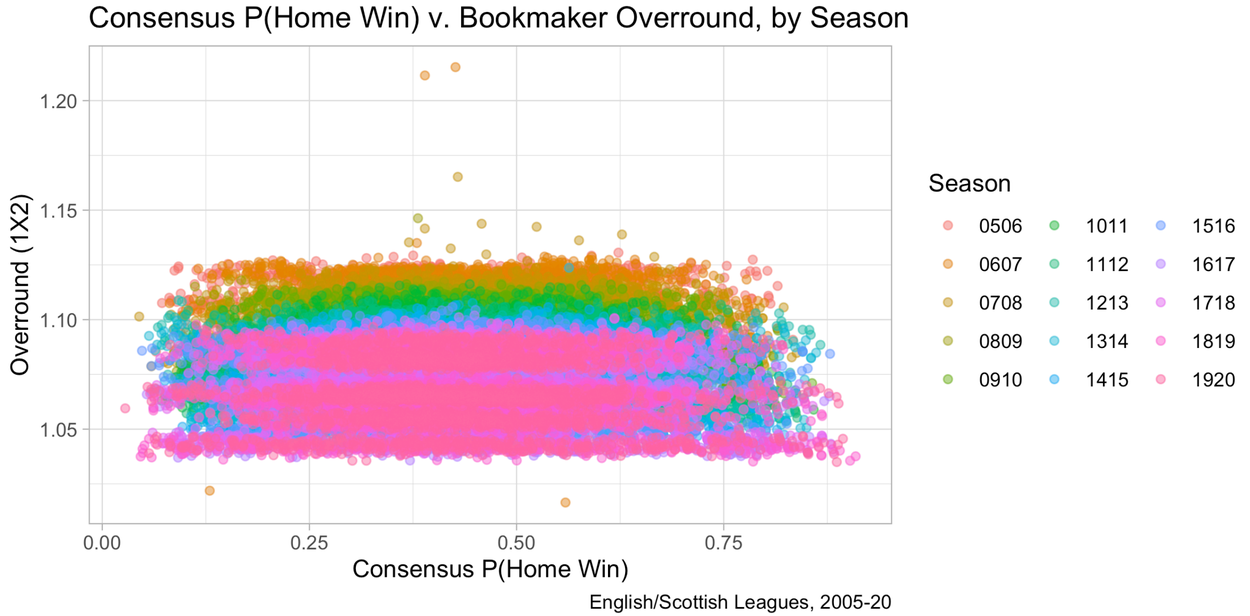
Level 3

Coventry University

Overround $\eta_i = \left(\sum \frac{1}{O_i}\right)$

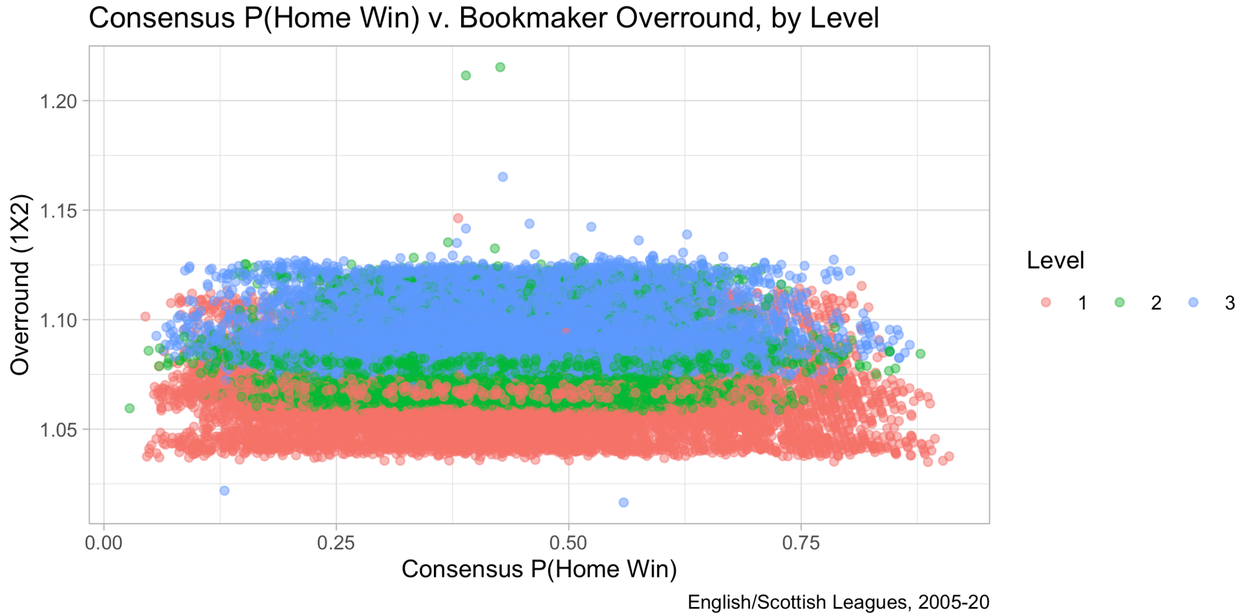
A brief look into where the bookmakers commissions lie

By SEASON



- Reducing over time for the 1X2 and UO markets
- AH overround is fairly constant at around 4.5%
- AH and UO market consensus probabilities are varying less

By LEVEL



- Level 1 is most **varied** in the 1X2 market, followed by 2 then 3 (with highest commission)
- Ranges from around **3%** to around **12%** commission

Coventry University

§4: A PROPOSED BETTING ALGORITHM

A proposed simple, applicable algorithm to place bets, based on our findings.

Coventry University

Our criteria for **placing bets**

Bets are placed on the **large favourites** across our most accurate markets:
The 1X2 Home/Away Wins markets; AH Home/Away Wins markets.

$$\begin{cases} & p_{mi} & < \mu_m + 0.5\sigma_m & B_{mi} = 0 \\ \mu_m + 0.5\sigma_m \leq & p_{mi} & < \mu_m + \sigma_m & B_{mi} = 1 \\ \mu_m + \sigma_m \leq & p_{mi} & < \mu_m + 1.5\sigma_m & B_{mi} = 2 \\ \mu_m + 1.5\sigma_m \leq & p_{mi} & & B_{mi} = 3 \end{cases}$$

σ_m
The overall std dev for the m^{th} market

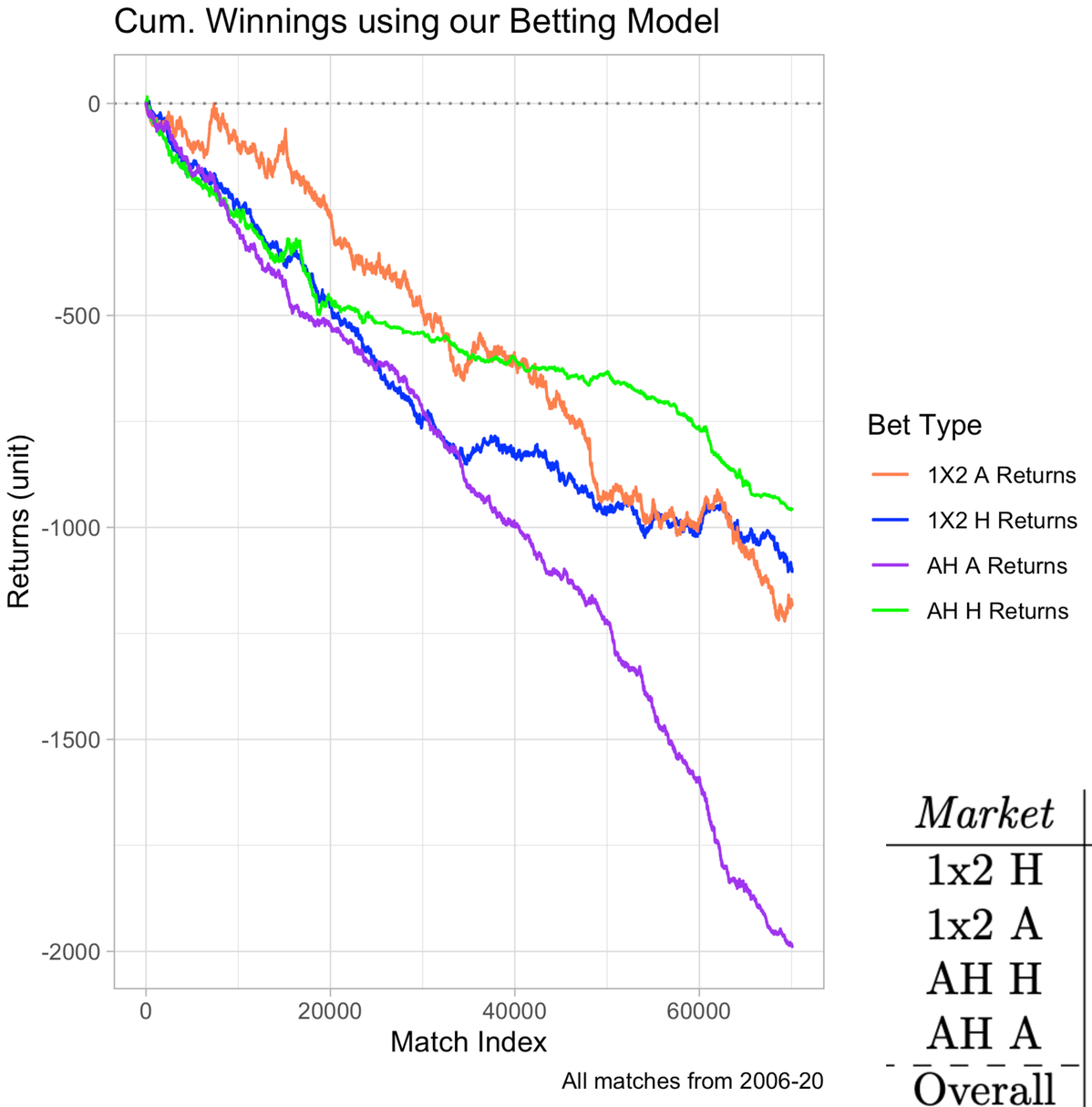
p_{mi}
The i^{th} match probability in the m^{th} market

μ_m
The overall mean for the m^{th} market

B_{mi}
The i^{th} match bet we place in the m^{th} market

Coventry University

RESULTS



- Despite high % accuracy of bets won (>50%) , a 4.3% **loss** was made on an 'investment' of 71,144 units
- For reference, Kaunitz et al. (2017) make a 3.5% profit with 44.4% of bets won
- **Asian Handicap Away** bets have the highest loss at −7.5%

Market	*Bets*	*Stake*	*Winnings*	*Proportion*	*Accuracy (%)*
1x2 H	19734	36923	–1103.380	–0.030	63.363
1x2 A	20018	37577	–1184.610	–0.032	47.472
AH H	11738	19607	–956.325	–0.049	45.604
AH A	19654	26353	–1988.525	–0.075	43.625
Overall	71144	120460	–5232.84	–0.043	50.509

Coventry University

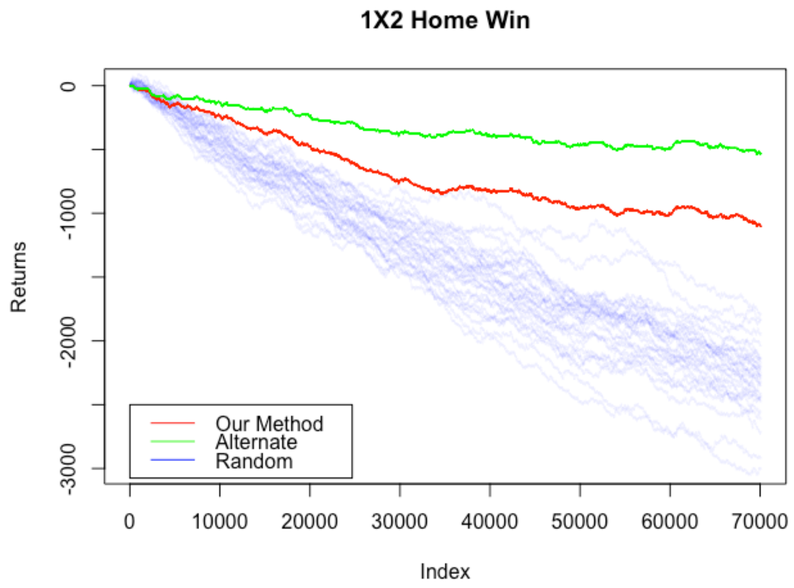
COMPARISON

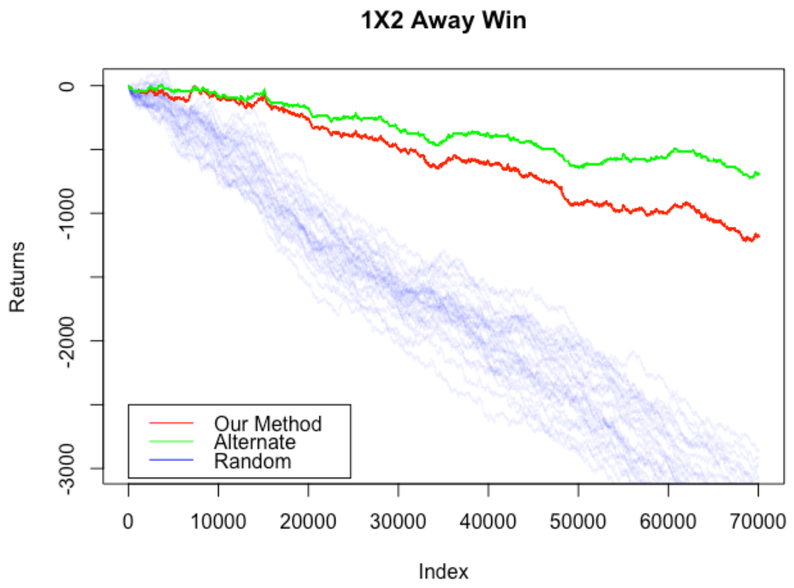
- This is compared against 10 runs of a **random bet strategy**, where a simulated "bettor" places bets with the same probability as our model.
- Our algorithm's accuracy (%) is far greater and lost less than the RBS.
- Much greater accuracy ≠ Much greater winnings

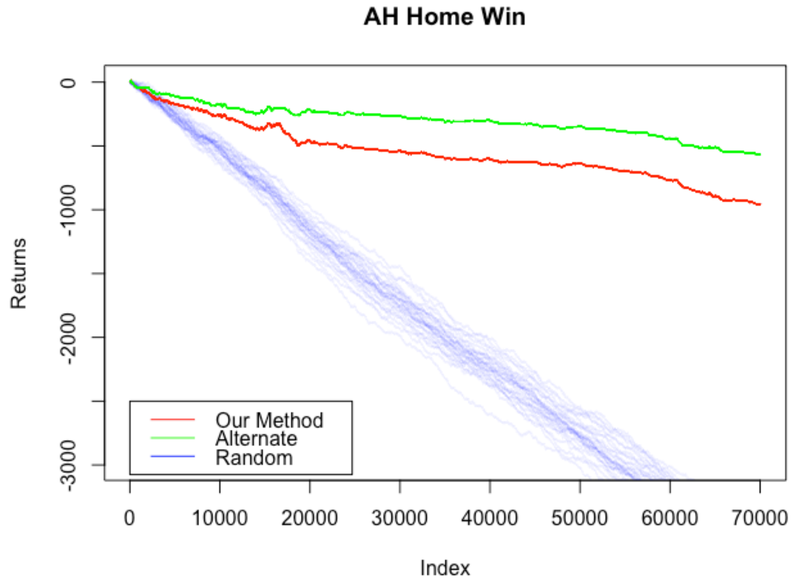
Market	*Bets*	*Stake*	*Winnings*	*Proportion*	*Accuracy (%)*
1x2 H	19652.1	39431.8	–2288.91	–0.058 (–0.028)	44.231 (–19.131)
1x2 A	20012.6	40284.6	–3182.22	–0.079 (–0.047)	30.038 (–17.434)
AH H	11723.3	20986.9	–3734.98	–0.178 (–0.129)	41.483 (–4.121)
AH A	19700.6	28330.5	–5018.98	–0.177 (–0.102)	40.319 (–3.305)
Overall	71088.6	129033.8	–14225.09	–0.110 (–0.067)	38.699 (–11.810)

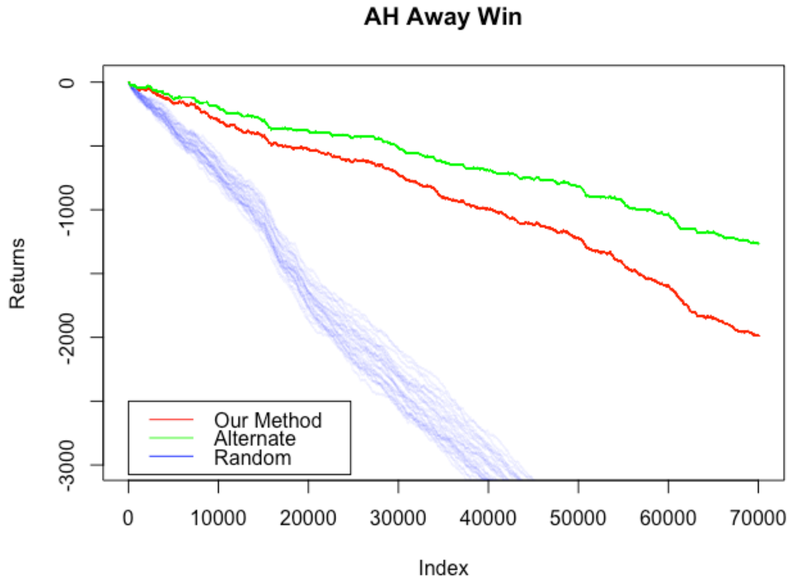
The differences are against our method

Coventry University









30 runs of the **random bet strategy** (blue) vs. **our algorithm** (red) vs. our ***alternate*** method (no bets on poor performing leagues) (green)

The alternate method has the same system as before, but doesn't bet on the German Bundesliga, French Ligue Une, English Leagues One and Two, and Scottish Championship

Coventry University

§5. CONCLUSION

Coventry University

Our Findings

- Betting odds offered are **very** accurate for Home Win & Away Win bets in the 1X2 and AH markets
- Bookmakers **struggle** to place accurate odds on **draws** and in the UO market
- Bookmaker commission has **reduced** over time
- **Competitive Balance** leads bookmakers to have more difficulty when placing odds
- Even a high accuracy when placing bets ≠ making winnings
- It is extremely hard to make **profit** without **high expertise** or **use of higher level tools** such as taking advantage of arbitrage opportunities or considering large numbers of bookmakers as Kauntiz et al. (2017)

Areas for possible further research:

- *Does the **style of play** (attacking or defensive) affect bookmaker accuracy?*
- *To what extent does **home advantage** affect bookmaker accuracy?*

Coventry University

What have I learnt?

- Use of R to solve applicable problems using a large dataset
- Time management
- LaTeX write up
- Explaining findings
- Research, reviewing literature and critical analysis of other people's works
- Methods for organising, and working on, large scale projects

THANK YOU FOR LISTENING

ANY QUESTIONS?