Inefficiency Across Segments of the U.K. Football Market

Team 1

David Buch
Anuj Gupta
Julian LaNeve
Dmitry Silantyev

March 2021

Contents

1 Executive Summary

With annual returns over of €10 billion and a thriving research literature searching for biased odds in sports betting, one might not expect to find inefficiency in the football betting market. However, the advent of data driven approaches provides means to not only test existing hypotheses but also analyse long term paradigm shifts in football markets. In this report, we provide an in-depth analysis of inefficiencies across different segments of the UK football economy: player contracts, gambling and odds-setting, and publicly traded financial assets. Broadly, we identify and test hypotheses to explain arbitrage opportunities, team management errors, and options trading anomalies. A brief overview is as follows:

- To identify bias in odds setting, we need an estimate of the ground truth - a reliable way to identify favorites and quantify uncertainty. To do so, we create a version of the Bradley-Terry model to determine "ELO" scores for each team's offense and defense. We find this strategy often beats market book makers, and provides interpretable insights into team heterogeneity.
- To be able to identify persistence of arbitrage opportunities in the betting market, we need to explain why different time periods allow for different levels of arbitrage. To that end, we use a first order regime-switching Markov regression model to statistically segment the data and identify two distinct regimes, explaining conflicting historical literature conclusions on arbitrage availability.
- To understand market inefficiency in the sports betting market, we create synthetic bookmakers and utilize risk-weighted bet sizing methods to identify arbitrage opportunities and potential profit. We then fit several regression models with various predictors on both the presence and degree to which arbitrage was available, building a coherent theory of what makes arbitrage possible.
- To understand if having an edge in predicting a match outcome can be translated into superior returns for investors, we relate the level of information entropy in the betting market to idiosyncratic volatility of stock returns and dispel the earlier findings of statistically significant relationship between the two.
- To confirm the hypothesis of whether option markets can be relied upon in providing a better implied distribution of match results, we reconstruct an implied risk-neutral density from observed option prices and find that it is likely under-pricing the right tail of the distribution of stock returns. We suggest a simple option-based strategy that might help investors take advantage of current market conditions.
- To understand market inefficiency in the player markets, we explore various trends in the "value" of a player. We find that, on average, older offensive soccer players are typically undervalued compared to younger offensive soccer players.

This led to an innovative model for team ratings that extends existing methods and outperforms bookmakers in setting match outcome odds. We use this model - which can be explored further in our **R Shiny App** (https://davidabuch.shinyapps.io/uk_football_analysis/) - to reach key insights about betting market inefficiency:

- Greater disagreement among bookies results in both higher arbitrage likelihood and more profitable arbitrage bets
- Arbitrage, although still somewhat available, is disappearing at unprecedented rates due to increased public awareness and a drop in the number of bookmakers operating in the market
- Football matches with uneven team skills tend to result in higher arbitrage likelihood
- Home team advantage disappeared in 2020-2021 season due to lack of fan attendance

We finally suggest avenues for further research in areas of player skill, bet sizing, and option trading viability.

2 Technical Exposition

2.1 Offense/Defense ELO: A Novel Method for Team Quality Evaluation and Match Outcome Forecasting

The outcomes of professional football matches, like those of many other sports, are understood to derive from complex interactions of play styles, multifaceted skill profiles, and match circumstances [25]. In spite of this, the countless quantitative team ratings systems promulgated in the academic and sports entertainment literature distill team ratings to a single univariate quantity [17, 1]. Worse yet, many popular techniques cannot account for match scores or opponent quality, instead relying exclusively on ternary 'Win-Loss-Draw' information while at best deploying *ad hoc* rules to either keep or entirely disregard data from 'noncompetitive' match-ups [10].

The widespread public availability of quantitative match outcome data, paired with the rapid pace of statistical methodology development over the last 30 years renders the kludge-like nature of the most popular sports rating systems somewhat startling. To address this gap, we have devised a model for team rating that eschews those pitfalls by leveraging the interpretability of the ELO system [13], and the statistical coherency of the Bradley-Terry model [5]. However, using modern statistical techniques it expands upon those methods to **accommodate detailed match score outcomes** and enable learning of **both *offensive* and *defensive* dimensions of team skill**.

We define our Offense/Defense ELO model as follows: Conditional on the expected score for both the home and away teams in our expected match-ups, we treat the observed scores as independent Poisson draws as suggested by discussions in Dixon and Coles, 1997 and Cain *et al.*, 2000 [9, 7]. We characterize a match via an ordered pair (H_g, A_g) indicating the home and away teams that participated in match g. For ease of notation, we drop the subscript g and assume the following model statement is iterated independently across all matches g

$$\begin{aligned} Y_H \sim \text{Pois}(\mu_H) \quad Y_A \sim \text{Pois}(\mu_A) \\ \mu_H = \exp(\alpha_0 + O_H - D_A) \\ \mu_A = \exp(\alpha_1 + O_A - D_H) \end{aligned}$$

where each team i is associated with a two-dimensional latent parameter (O_i, D_i) reflecting their offensive and defensive skill, while α_0 and α_1 are offset terms that capture the log expected number of points scored by the home and away team, respectively, in a match up of evenly skilled teams. These offsets are assumed to be fixed for a particular league and season, though they allowed us to replicate a collapse in home team advantage during the COVID-19 lockdowns that had been previously documented in the literature [24].

The key insight to gain from the model equation above is that our method bears a great resemblance to classical Poisson regression, but has been adapted to accommodate data drawn from networks of head-to-head comparisons. The very spare model specification above can be completed by placing priors on the team skill parameters, analogous to a random-intercept Poisson regression. To make this explicit, for all teams i we have

$$\begin{aligned} (O_i, D_i) \sim N(0, \Sigma) \\ \Sigma \sim \text{Inv-Wish}(\nu, I) \\ \alpha_j \sim N(0, 10), \end{aligned}$$

where Σ is the unknown covariance of the latent skill parameters. The variance of the offset prior distributions can be specified in advance because these parameters have an easily-traced impact on match final scores: an α_0 greater than 10 would imply that in an evenly matched football game, the home team would be expected to score 22,000 points! Clearly, then, our prior is sufficiently uninformative.

Alternatively, more insights can be drawn from the ELO model by conducting analysis conditionally, assuming

$$O_i = X_i \beta_0$$
$$D_i = X_i \beta_D$$

where X_i are team covariates, such as player salaries.

We were able to fit the ELO model via MCMC techniques despite its complex, non-conjugate construction and large dimensionality (in the random effects case), by making use of highly efficient Hamiltonian Monte Carlo techniques [19] via the user friendly open-source library Stan [22].

For a richer introduction to the potential of this model, refer to our accompanying **R Shiny App** (https://davidabuch.shinyapps.io/uk_football_analysis/). Note that, for ease of interpretation, where the latent skill parameters are presented on their own, we scale them up by a factor of 100 and describe them as "Offence-Defense ELO Scores."

Given public appetite for team ratings/rankings information across all sports, as well as the remunerative opportunities derived from sophisticated team evaluation and reliable match outcome forecasting with uncertainty quantification, we see this model as possessing inherent value to sports team managers, sports gambling bookmakers, and sports entertainment media services. For this report, however, in keeping with our theme of identifying and quantifying instances of inefficiency in UK football markets, we focus on just two applications of our team rating model: football match odds-setting, and investigating the relationship between team composition and team quality to discover biases in player valuation.

In Table 1, we use the standard Ranked Probability Score measure (see Equation 2.1; r is the number of potential outcomes, p_j and e_j are the probability forecasts and observed outcomes at position j) [8] to rate the odds-setting performance of our model against the published book odds of eleven popular bookmakers, and see that our model dominates the group. Though it is important to consider that book makers account for factors beyond match outcome probability, this justifies our use of our models odds as a ground truth in analyzing arbitrage and biases in the sequel.

$$RPS = \frac{1}{r-1} \sum_{i=1}^{r-1} \left(\sum_{j=1}^{1} (p_j - e_j) \right) \tag{2.1}$$

Bookie Name	RPS
Model	0.206884
SB	0.212571
BS	0.212633
VC	0.212842
B365	0.212863
LB	0.213005
WH	0.213083
SJ	0.213114
PS	0.213185
BW	0.213265
IW	0.213440
GB	0.213459

Table 1: RPS Scores of Bookies and Novel Offense/Defense ELO Model

2.2 Time Clustering through Markov Regime Model

An analysis of prior literature on arbitrage in European sports betting markets demonstrates that researchers come to disparate conclusions with regards to the prevalence of arbitrage. Vlastakis *et al.* find that 0.5% of all matches offer arbitrage opportunities using European soccer match data from 2002-2004 across 6 bookies [23]. Franck *et al.* find that almost 4% of all matches offer arbitrage opportunities using betting odds from 10 bookies from 2004 - 2011 [11]. Grant *et al.* find arbitrage available in 25.6% of matches from six major European soccer leagues across six bookies from 2011 - 2013 [14]. Arbitrage availability post-2013 has not been addressed in popular literature.

Clearly, the availability of arbitrage in European soccer matches is inconsistent across time. In order to address this in our analysis, we utilize a Markov regime switching regression model using arbitrage availability as a dependent variable to identify separate regimes, clustering time periods according to arbitrage behavior. This results in two separate regimes: periods of "low availability" of arbitrage, and periods of "high availability" of arbitrage. Table 2 and Figure 1 show the results of this model.

Regime 1	**coef**	**std err**	**z**	**P> |z|**	**[0.025**	**0.975]**
const	0.0081	0.004	2.237	0.025	0.001	0.015
Regime 2	**coef**	**std err**	**z**	**P> |z|**	**[0.025**	**0.975]**
const	0.0855	0.005	18.225	0.000	0.076	0.095
Non-Switching Params	**coef**	**std err**	**z**	**P> |z|**	**[0.025**	**0.975]**
sigma2	0.0001	4.66e-05	2.828	0.005	4.05e-05	0.000
Switching	**coef**	**std err**	**z**	**P> |z|**	**[0.025**	**0.975]**
p[0->0]	0.9181	0.084	10.972	0.000	0.754	1.082
p[1->0]	0.2046	0.153	1.334	0.182	-0.096	0.505

Table 2: Markov Regime Switching Model Results

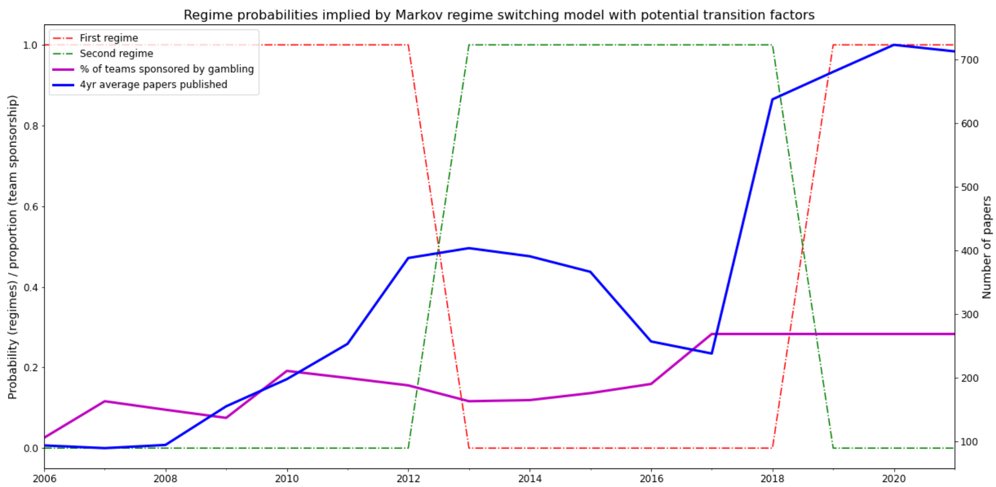

Figure 1: Markov Regime Switching Model Resulting Regimes with Switching Factors

We also display the proportion of teams sponsored by gambling companies and a 4 year backwards rolling average number of papers published on European football arbitrage in Figure 1. We utilize data from Bunn *et al.* (2019) on U.K. football team sponsorship to determine the proportion of teams sponsored by gambling companies [6]. Although not statistically rigorous, it appears the second regime transition may occur as a result of increased gambling sponsorship. This is additionally supported by van Huet (2017), who shows that sponsorship of Premier

League clubs by bookmakers had a major impact on arbitrage availability [16]. Since footballs fans are more likely to place bets on their favourite team with the bookmaker who sponsors it, individual bookmaker's books become imbalanced and in response they decrease the odds for the favorite no matter what they believe the true odds are, thus opening potential arbitrage opportunity.

Further, the transition from the second regime back to the first regime may be driven by an increase in academic literature on the topic. It is well known that increased awareness and exploitation of arbitrage generally decreases arbitrage availability, so when treating academic literature as a proxy for public knowledge, it follows that increased academic publications on European football arbitrage may reduce arbitrage availability.

For the purpose of our analysis, we separate the first and second instance of the "low availability" regime into two separate regimes to provide additional perspective into recent arbitrage activity in betting markets. Hence, we define the period from the beginning of the 2005/2006 season to the end of the 2012/2013 season "Regime 1," the period from the beginning of the 2013/2014 season to the end of the 2018/2019 season "Regime 2", and the period from the beginning of the 2019/2020 season onward "Regime 3."

2.3 Arbitrage Opportunity Behavior Across Regimes

To identify the presence and degree of arbitrage availability in our data, we begin by creating a synthetic bookie using the maximum available odds across all 11 bookies. This synthetic bookie is representative of the maximum profit we can make. Accordingly, if we can allocate bets to the synthetic bookie such that we make money on any outcome, we know arbitrage is possible; otherwise, arbitrage is not possible.

We simulate returns of a portfolio investing solely in arbitrage opportunities using a total of $100 per match, spread across all odds. We use a uniform allocation method guaranteeing the same profit regardless of the outcome in accordance with Vlastakis *et al.* (2009) with weights

$$\hat{w}_i = \frac{P_i}{\sum_{i=1}^{n} P_i} = \frac{\frac{i}{b_i}}{\sum_{i=1}^{n} \frac{i}{b_i}} \tag{2.2}$$

placed on each i^{th} outcome over n possible outcomes, where P_i is the probability of the i^{th} outcome implied by the bookie using the relationship

$$P_i = \frac{1}{b_i} \tag{2.3}$$

and b_i is the bookie's odds of the i^{th} outcome [23]. Note that each b_i and P_i comes from the synthetic bookie, created using the maximum odds for each outcome i across all bookies. The result of $100 bets placed in each arbitrage opportunities is shown in Figure 2 below. Our returns are consistent with those found in previously mentioned literature but provide additional insight past 2013. Generally, as the frequency of football matches with arbitrage increases, the strategy's cumulative returns increases.

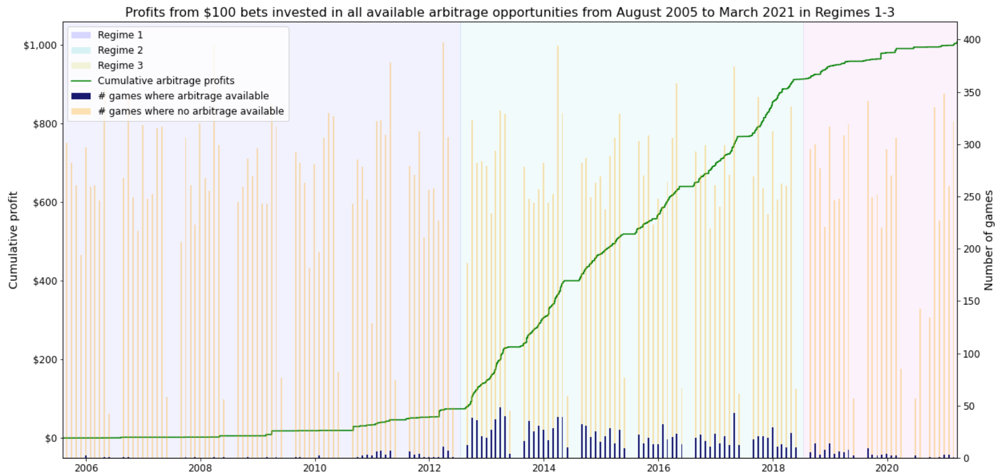

Figure 2: Arbitrage Bet Returns from 2006 - Present

From the 2005/2006 football season to present, the strategy returns total over $1,000 (excluding basis) with a basis of $100 (1,000% return) across 1,525 matches. This averages to an approximate 0.65% return per game and 17% per year, which is extremely significant given the risk-free nature of the bets. However, we admit that bookies typically charge fees and certain bookies only allow bets to be placed in-person, which may decrease the availability and profitability of our identified arbitrage opportunities.

2.3.1 Arbitrage availability across and within regimes

To better understand how our results of arbitrage availability relate to those of Vlastakis (0.5% of all matches), Franck (4% of all matches), and Grant (25.6% of matches), we quantify the effect of regime and time within regime on arbitrage availability using a logistic regression. Vlastakis analyzed matches from pre-Regime 1, Franck analyzed matches from Regime 1, and Grant analyzed matches from Regime 2. Prior to statistical analysis, we hypothesized Regime 2 would offer significantly more arbitrage opportunities based on the disparate results from Franck and Grant. We fit a logistic regression of the form

$$\begin{aligned} \ln\left(\frac{P_{arb}}{1-P_{arb}}\right) &= \beta_0 + \beta_1 * R_2 + \beta_2 * R_3 \\ &+ \beta_3 * Dur + \beta_4 * Dur * R_2 + \beta_5 * Dur * R_3 \end{aligned} \tag{2.4}$$

where P_{arb} indicates the probability of arbitrage, R_i is an indicator variable for the i^{th} regime, and Dur is the number of years after the beginning of the current regime. The results of our logistic regression model are shown in Table 3.

	coef	std err	z	P> |z|	[0.025	0.975]
Intercept	-4.7031	0.086	-54.408	0.000	-4.873	-4.534
R_2	5.8437	0.089	65.940	0.000	5.670	6.017
R_3	4.4875	0.094	47.514	0.000	4.302	4.673
Dur	0.5772	0.016	35.669	0.000	0.546	0.609
Dur * R_2	-0.6810	0.017	-39.595	0.000	-0.715	-0.647
Dur * R_3	-1.1250	0.033	-34.202	0.000	-1.189	-1.061

Table 3: Logistic Regression Results on Arbitrage Availability Indicator

The model has an overall p-value of < 0.001 and predicts arbitrage availability with 75% accuracy, giving us confidence in its results. The model confirms our hypothesis: Regime 1 has significantly less arbitrage than Regimes 2 and 3. This also confirms the results of Vlastakis, Franck, and Grant, and helps explain the disparate results. Further, the duration in regime covariate sheds additional light on arbitrage availability within each regime. Regime 1 has generally increasing arbitrage availability, whereas Regimes 2 and 3 have generally decreasing arbitrage availability. Further, the degree to which arbitrage is decreasing within Regime 3 is significantly more than Regime 2, indicating that arbitrage is currently disappearing quickly.

2.3.2 Arbitrage availability in even and uneven matches

After understanding how arbitrage changes relating to regimes, we wanted to explore the relationship between arbitrage and team match-ups. We hypothesize that games where teams are more evenly matched (taking into account home team advantages) will result in higher likelihood of arbitrage. Intuitively, we expect that games that are more evenly matched are inherently less predictable and thus bookies should have a tougher time setting correct odds. To test this, we use the ELO-type team scores described in Section 2.1 to determine the expected probability of the home and away team winnings, and define the win chance spread as the difference in probabilities. We then create a model of the form

$$\begin{aligned} \ln\left(\frac{P_{arb}}{1-P_{arb}}\right) &= \beta_0 + \beta_1 * R_2 + \beta_2 * R_3 \\ &+ \beta_3 * Spread + \beta_4 * Spread * R_2 + \beta_5 * Spread * R_3 \end{aligned} \tag{2.5}$$

where P_{arb} indicates the probability of arbitrage, R_i is an indicator variable for the i^{th} regime, and *Spread* is the absolute difference between the implied probability of the home and away team winning from the Bradley-Terry-like model described in 2.1. The results of our logistic regression model are shown in Table 4.

	coef	**std err**	**z**	**P> |z|**	**[0.025**	**0.975]**
Intercept	-2.3697	0.040	-59.271	0.000	-2.448	-2.291
R_2	2.9003	0.043	67.090	0.000	2.816	2.985
R_3	0.6750	0.058	11.668	0.000	0.562	0.788
Spread	0.0096	0.001	7.345	0.000	0.007	0.012
Spread * R_2	0.0039	0.001	2.758	0.006	0.001	0.007
Spread * R_2	0.0226	0.002	12.536	0.000	0.019	0.026

Table 4: Win Chance Spread Model Results

The model has an overall p-value of < 0.001 and predicts arbitrage availability with 63% accuracy, giving us confidence in its results. The results of the model contradict our hypothesis, but are significant nonetheless. Rather, the data illustrates that as teams are less evenly matched (i.e., there is a clear expected winner), the likelihood of arbitrage increases. Further, in Regime 3, this effect is even more extreme, indicating that arbitrage is presently more prevalent in uneven matches than in the past.

2.3.3 Arbitrage availability resulting from bookie disagreement

Our final hypothesis related to arbitrage availability relates to bookie disagreement. We hypothesize that if bookies disagree about the likelihood of an outcome, they are more likely to set dissimilar odds, resulting in a higher probability the maximum of all odds gives an opportunity for arbitrage. To test this hypothesis, we use the standard deviation of the implied probabilities of each bookies' outcomes as a proxy for general bookie disagreement. Figure 3 graphically shows the change in bookie disagreement across regimes.

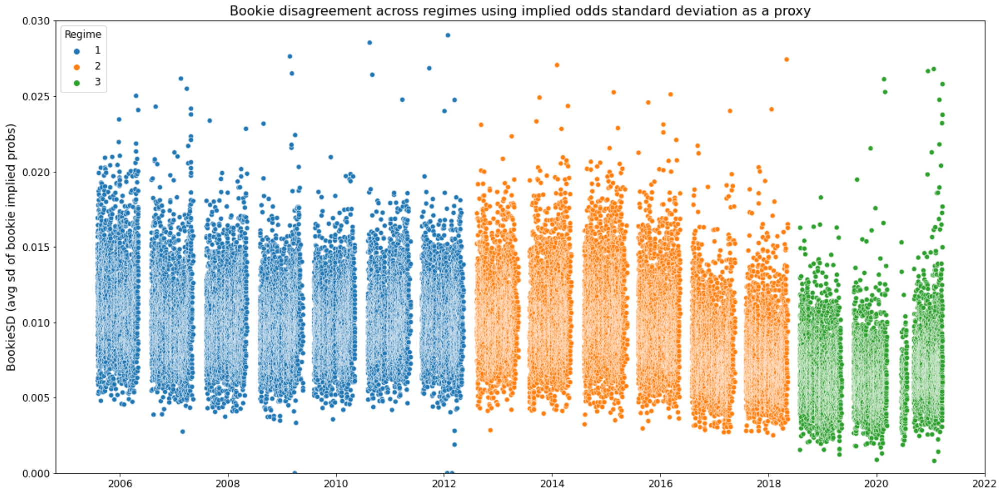

Figure 3: Bookie Disagreement Across Regimes

We then create a logistic regression model to statistically validate our hypothesis of the form

$$\ln\left(\frac{P_{arb}}{1-P_{arb}}\right) = \beta_0 + \beta_1 * R_2 + \beta_2 * R_3 + \beta_3 * SD + \beta_4 * SD * R_2 + \beta_5 * SD * R_3 \quad (2.6)$$

where P_{arb} indicates the probability of arbitrage, R_i is an indicator variable for the i^{th} regime, and SD is the average standard deviation of probabilities across outcomes implied by bookie odds. The results of this model are shown in Table 5.

	coef	std err	z	P> |z|	[0.025	0.975]
Intercept	-7.7338	0.129	-59.774	0.000	-7.987	-7.480
Regime[T.2]	3.4961	0.140	25.017	0.000	3.222	3.770
Regime[T.3]	-0.1030	0.205	-0.502	0.615	-0.505	0.299
BookieSD	453.4318	9.548	47.488	0.000	434.717	472.146
BookieSD:Regime[T.2]	10.9060	10.756	1.014	0.311	-10.176	31.988
BookieSD:Regime[T.3]	331.0699	19.528	16.954	0.000	292.796	369.344

Table 5: Bookie Disagreement Model Results

The model has an overall p-value of < 0.001 and predicts arbitrage availability with 83% accuracy, giving us confidence in its results. The model results confirm our hypothesis that arbitrage is generally more available as bookies disagree. Further, this effect changes across regimes. Regime 1 and Regime 2 seem to have a similar relationship between bookie disagreement and prevalence of arbitrage, but the Regime 3 interaction demonstrates a much more extreme relationship; bookie disagreement in Regime 3 is more significant in arbitrage prevalence than it is in Regimes 1 and 2.

We then hypothesize that increased bookie disagreement should result in more profitable arbitrage opportunities where arbitrage exists. We continue to use standard deviation of bookies' implied probabilities as a proxy for bookie disagreement. However, instead of using the presence of arbitrage as the dependent variable, we instead use the profit made from each arbitrage bet. We create an ordinary least squares linear regression model of the form

$$profit_{arb} = \beta_0 + \beta_1 * R_2 + \beta_2 * R_3 + \beta_3 * SD + \beta_4 * SD * R_2 + \beta_5 * SD * R_3 \quad (2.7)$$

where $profit_{arb}$ is the profit on a \$100 bet spread across odds, R_i is an indicator variable for the i^{th} regime, and SD is the average standard deviation of probabilities across outcomes implied by bookie odds. The results can be found in Table 6.

	coef	**std err**	**t**	**P> |t|**	**[0.025**	**0.975]**
Intercept	-1.3255	0.333	-3.978	0.000	-1.979	-0.672
R_2	0.4021	0.347	1.158	0.247	-0.279	1.083
R_3	-0.0859	0.397	-0.217	0.829	-0.864	0.692
SD	143.6091	20.820	6.898	0.000	102.770	184.448
SD * R_2	-23.5546	22.068	-1.067	0.286	-66.842	19.733
SD * R_3	30.1560	25.978	1.161	0.246	-20.801	81.113

Table 6: Bookie Disagreement OLS Model Results

The results of the model reveals that more disagreement among bookies results in higher arbitrage profits. Unlike prevalence of arbitrage, this effect does not significantly change across regimes. As evidenced by the significance of the regime intercepts, the degree of bookie disagreement appears to explain the differing profits between Regimes 1, 2, and 3. Note that bookie disagreement does not, however, explain the differing prevalence of arbitrage across regimes.

2.4 Inefficiency in the Player Contract Market

A first insight, and an important lens through which the following analysis must be interpreted, is that football teams in the premier league are only as good as their budget allows. According to a regression of total OD-ELO versus total team salary, we find squad pay explains 50% of heterogeneity in team performance, so it is in the remaining half of variation that we explore why some teams under-perform relative to their payroll while others transcend it.

Due to the physicality involved in the sport, we set out to investigate the role of age on value (player performance divided by salary). We first created an ordinary least squares regression model of the form

$$\begin{aligned} v_{Player} &= \beta_0 + \beta_1 * T_{Neutral} + \beta_2 * T_{Offense} \\ &\quad + \beta_3 * Age + \beta_4 * Age * T_{Neutral} + \beta_5 * Age * T_{Offense} \end{aligned} \tag{2.8}$$

where v_{Player} is a player's individual skill (as measured by a third party) relative to their pay, T_i is a player's position, and Age is a player's age. The results of this regression are shown in Table 7. Across all players, we observe a clear negative trend between player value and age. This effect is most extreme for forwards ($T_{Offense}$), indicating that there may be a systematic mispricing of forwards as their age increases.

However, it is possible that these older forwards may benefit their teams in ways not quantified in the individual skill ratings data. It is plausible that, due to their greater experience, they are better able to cooperate with other players on the field and thereby justify their unexpectedly high price tags. To increase the strength of our evidence that older forwards are overpriced, we would need to evaluate whether squads relying on older forwards tend to underperform offensively relative to their total offensive spending. Now we make use of O-ELO estimates relative to forward squad spending to estimate the value of a complete forward squad, and fit an ordinary least square regression model of the form

$$v_{Offense} = \beta_0 + \beta_1 * Age \tag{2.9}$$

where $v_{Offense}$ forward squad value and Age is the squads average age. The results of this regression are shown in Table 8. Here, we again observe that as average age increases, value (performance relative to cost) declines.

There are, of course, alternative explanations for these observations that prevent us from making a definitive conclusion that these players are mispriced. Firstly, separate regression analyses revealed that, in general, the relationship between individual player skill and player salary is sublinear; in other words, clubs must pay a premium to hire generational talents (as opposed to 80^{th} percentile players). Since these top players are disproportionately older, the nonlinearity in skill versus salary could lead to false associations in our analysis of value versus age. Or, the mispricing of older players could be driving the sublinearity of skill versus pay. With the data provided, it is impossible to disentangle causality, but it is worth noting that the decline in value with age becomes less significant (-0.5, p-value 0.01 versus -1.2, p-value < 0.001) after adjusting for player salary, and the interaction effect with player category become nonsignificant (in other words, loss of value is not worse for forwards than defencemen, after adjusting for salary). Also, it should not go unmentioned that venerable stars may bring value to football clubs off the pitch: in 2018, Italian club Juventus reportedly sold $60 million in jerseys overnight after signing international superstar Cristiano Ronaldo [15]. Ultimately, even OD-ELO cannot capture the full richness of the football economy.

	coef	**std err**	**t**	**P> |t|**	**[0.025**	**0.975]**
Intercept	70.9906	6.828	10.397	0.000	57.593	84.388
$T_{Neutral}$	25.9699	11.193	2.320	0.021	4.008	47.932
$T_{Offense}$	22.9115	11.324	2.023	0.043	0.692	45.131
Age	-1.2606	0.263	-4.789	0.000	-1.777	-0.744
Age * $T_{Neutral}$	-1.2402	0.439	-2.823	0.005	-2.102	-0.378
Age * $T_{Offense}$	-1.3638	0.452	-3.017	0.003	-2.251	-0.477

Table 7: Effect of Player Age on Value Regression Results

	coef	**std err**	**t**	**P> |t|**	**[0.025**	**0.975]**
Intercept	3259.6622	793.990	4.105	0.001	1591.551	4927.773
Age	-113.2894	30.674	-3.693	0.002	-177.734	-48.845

Table 8: Effect of Age on Squad Offensive Value

2.5 Connecting the Dots: Relating Match Outcomes to Performance of Financial Assets

Having identified a prolonged period of exploitable inefficiencies in the betting market, we set out to explore a hypothesis of stock returns of publicly listed football clubs being affected by unexpected match outcomes. We investigate if having an edge in predicting a match outcome can be translated into superior returns for investors and whether option markets can be relied upon in providing a better implied distribution of match results.

Previous research on the topic produced somewhat mixed results: while earlier papers written in the 2000s generally discover statistically significant market reactions to unexpected match outcomes [3], more recent studies conclude that such variables as the number of spectators and transfers of players (that are both considered proxies for the financial performance of a club) explain most of the variation of stock returns [12]. Some attempts have been made to relate unexpected sporting results to periods of increased volatility [2] but none of them incorporate option market data, to the best of our knowledge. It is customary in the literature to use a cross-section of returns across all publicly traded football clubs to establish the strength of the connection between sporting events and financial markets. We would however argue that cross-sectional results do not offer any actionable conclusions for investors as averaging responses of different stocks across matches held over several years can not be translated into a viable trading strategy. Instead, we only analyze the shares of the largest publicly traded football club, Manchester United (ticker: MANU, re-listed on NYSE in 2012), as they provide sufficient market depth across betting, stock and option markets. While our analysis is much narrower in scope, we attempt to offer deeper insights into the relationship between the three market segments. Liquidity is an important factor as any discovered market inefficiency can not be exploited in the absence of sufficient market depth. Having analyzed stock returns of other football clubs, most notably Juventus and AS Roma, we admit that their distributions exhibit even heavier tails, which suggests more potential trading opportunities on paper, but likely less real use for any institutional investor, particularly given how shallow their option markets are.

We begin by estimating a CAPM regression of daily MANU returns on market returns in order to separate market-wide shocks from idiosyncratic shocks related to MANU's performance (see Equation 2.10). Despite being listed on NYSE, MANU doesn't exhibit strong correlation with the broader US stock market and is more attuned to the changes in the FTSE Index. We convert the latter into USD to ensure that we deal with dollar returns in both cases.

$$R_{MANU} = \alpha + \beta R_{FTSE} + \epsilon \tag{2.10}$$

This simple exercise already uncovers several interesting properties of MANU returns. The R^2 of just 7.7% indicates that only a small portion of variation of returns can be attributed to the overall market performance. Yet, $\hat{\beta} = 0.447$ has a standard error of only 0.033, pointing to high statistical significance of the relationship. This is somewhat contrary to our expectation of pro-cyclical football clubs' stocks being high-beta. Studying the distribution of residuals also reveals that it has excess kurtosis of 5.5 and that the proportion of 3-sigma and larger events exceeds 1.15% of the sample, comparing to 0.27% for the Normal distribution. While high peakedness of the distribution does not offer any actionable opportunities for market participants, the fat tails do and we are going to pursue this idea further.

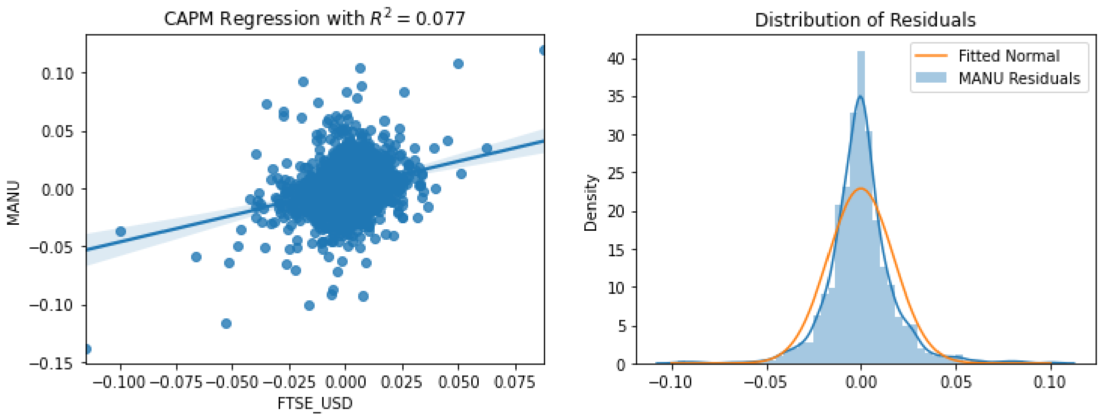

Figure 4: CAPM Regression and Distribution of Residuals

Given that more than 77% of all matches in our 2012-2021 sample was played on weekend, and in line with previous literature, we focus on excess stock returns on trading days following matches and present aggregated results in Table 9. The overall directions of the reactions are in line with our intuitions, except possibly in the case of a win in an away match, and with previous research findings which showed that the stock market is much more sensitive to defeats and draws than to wins [20]. In general, while the sample's excess kurtosis increases to 8.55, indicating even fatter tails, otherwise the sample of residuals on trading days following sporting event matches the overall sample distribution almost percentile-by-percentile. That and the magnitudes of the standard deviations make us reject the hypothesis of unexpected match results offering any specific trading opportunities.

	Home		Away	
	Mean	St. Dev.	Mean	St. Dev.
Win	0.21%	1.62%	-0.14%	2.13%
Draw	-0.20%	1.51%	-0.32%	1.19%
Defeat	-0.38%	1.55%	-0.10%	1.55%

Table 9: Excess Return on Trading Days Following Matches

One of the reasons behind the high dispersion of excess returns following similar events might lie in misalignment between investors' and bettors' expectations regarding match outcomes [4]. While it is not straightforward to infer investors' expectations about match outcomes, we assume that the higher uncertainty about them would be reflected in higher volatility of stock returns. On the other hand, bettors' expectations can be inferred directly from the betting odds that are calibrated against both realized probabilities and bettors' demand. We first plot rolling realized probabilities of winning and losing a match against the probabilities implied by the betting odds in Figure 5. In our view, this dynamic suggests that while in general implied probabilities are well calibrated, there are periods spanning several seasons when the uncertainty about match outcomes is very high. To combine these metrics into one, we proxy bettor's uncertainty by Shannon's entropy in Equation 2.11 and plot it in Figure 6.

$$H(X) = -\sum_{i=1}^{n} p(x_i) \log p(x_i) \tag{2.11}$$

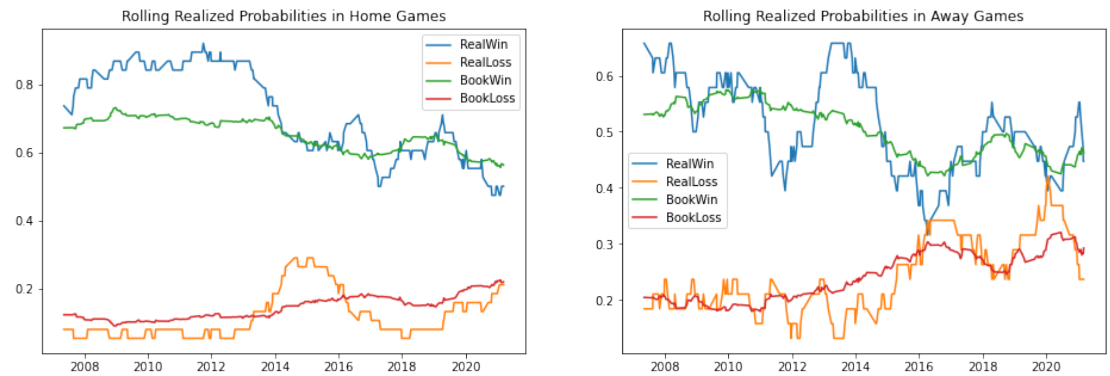

Figure 5: Implied vs. Realized Probabilities of Match Outcomes

Here, again our hypothesis is not confirmed by the data: while we clearly observe two different entropy regimes (a period of rising uncertainty in 2014-2016 followed by a sharp decrease in 2017 and an increase in 2018-2021), MANU's idiosyncratic volatility, defined as rolling standard deviation of CAPM residuals, does not exhibit any visible regime change in 2012-2021. Our results so far indicate that whatever edge an investor might have in predicting match outcomes, it is unlikely to result in any sustainable trading strategy. Still, we believe that unusually heavy tails of the distribution of residuals might be exploitable.

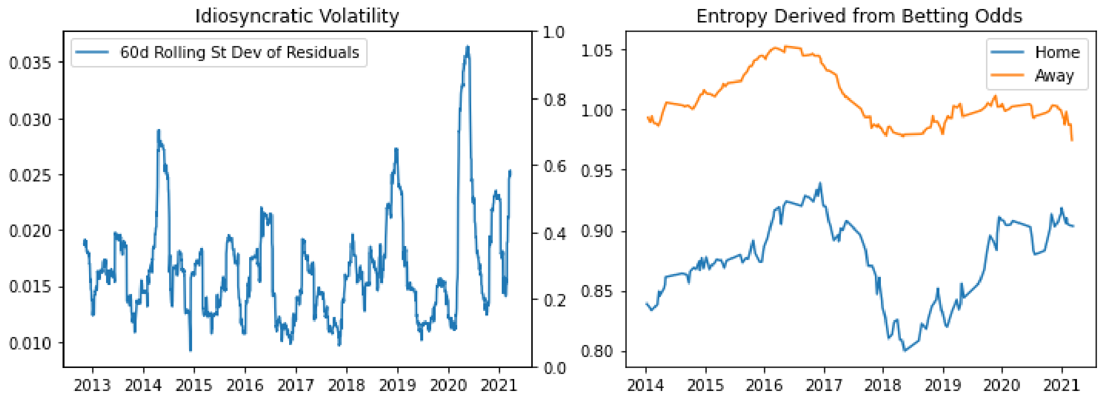

Figure 6: Idiosyncratic Volatility Across Two Entropy Regimes

In order to test whether option markets price in the extreme volatility in both tails, we employ the seminal Shimko methodology of inferring a risk-neutral density of stock prices at maturity of the options from observed option prices [21]. We use option prices for 200-day call and put options on MANU as of 30 November 2020 in order to illustrate the method. All options expire on 18 June 2021 and thus should capture a wide range of outcomes until the end of 2020/2021 season. To ensure that the resulting density is continuous, we first need to estimate an implied volatility curve by using Weighted Least Squares to fit a quadratic polynomial from Equation 2.12. The next step is to use the relationship between the risk-neutral density $q_t(x)$ and the second derivative of call option price with respect to strike from Equation 2.13. We present the resulting implied volatility curve and the risk-neutral density in Figure 7.

$$\hat{\sigma}(K,t) = a(t)K^2 + b(t)K + c(t) \tag{2.12}$$

$$q_t(x) = e^{rt}\frac{\partial^2 c(K,\sigma(K))}{\partial K^2}\bigg|_{K=x} \tag{2.13}$$

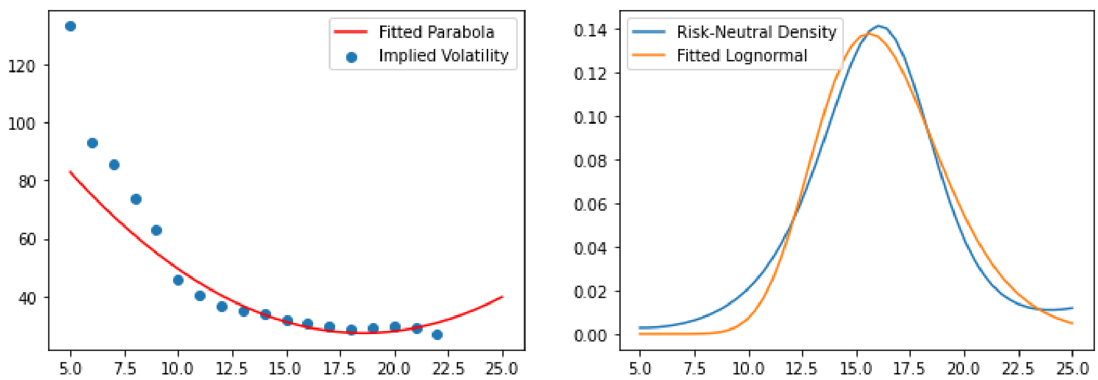

Figure 7: Implied Volatility Curve and Inferred Risk-Neutral Density

The inferred density is non-parametric in nature, can not be converted into a corresponding normal distribution by taking a logarithm and is only compared to log-normal distribution because the latter is the benchmark model for stock prices in quantitative finance. We still believe that examining its properties is illuminating because it exhibits a negative skew and an excess kurtosis of only 0.53. Furthermore, it is clear that a parabola does not provide a good fit for the observed structure of implied volatility, underestimating volatility at lower strikes and overestimating volatility at higher strikes. The implication is that the risk-neutral density should have had an even heavier left tail and a much smaller right tail. This leads us to conclude that the asymmetry in option pricing commonly referred to as "volatility smile" does not fully capture the tendency of MANU to jump upwards as wildly as downwards. While the premise of volatility smile is based on empirical observation that stocks generally fall faster than they increase in price, this is not the case with MANU, as we have seen earlier in the report.

In our view, this inefficiency can be exploited by a simple risk reversal strategy that would entail a delta-neutral combination of bought high-strike calls and sold low-strike puts, with the addition of dynamic delta-hedging whenever the position moves out of balance. At current prices, in order to construct this portfolio with no initial investment, an investor would only need to sell low-strike puts with a fifth of the notional of the bought calls, which further limits downside potential of the strategy. Limitations on the availability of public data on historical option prices do not allow us to backtest this strategy but we believe that these findings warrant further research into the topic.

3 Suggestions for Future Research

With the data we have already collected and cleaned, and using the robust methods developed in the above analyses, there are several lines of research we could pursue moving forward. A few of the most promising are:

- We have determined that older players, particularly forwards, are often overpaid relative to their skill rating and even their in-game performance. However, it is known that in some instances [15], esteemed star players can contribute to team revenue through buoyed ticket and merchandise sales. Can we quantify the financial impact of 'venerable elder' star players beyond their performance on the pitch?
- To determine whether there is an interaction effect between age and match time in determining team performance. Do younger teams underperform/older teams outperform their expected skill level in daytime matches?
- We have clearly identified the existence of small biases among bookmakers in their odds setting process. However, it is not clear that these biases are substantial enough to allow an unbiased predictor to win money against bookmakers after margin is taken into account. Though the likelihood of opportunities increases if we are allowed to consider dynamic evolution of price changes. An iterative application of our model could be backtested with a optimal (Kelly Criterion) bet placement strategy to investigate whether there are financially rewarding opportunities in the football betting market.
- A more robust method [18] of recovering risk-neutral density from option prices can be applied to better quantify the degree to which the right tail of the distribution of football clubs' stock returns is underpriced in the option market.

References

[1] Barrow, Daniel, et al. ”Ranking rankings: an empirical comparison of the predictive power of sports ranking methods.” Journal of Quantitative Analysis in Sports 9.2 (2013): 187-202.

[2] Benkraiem, Ramzi, Frédéric Le Roy, and Waël Louhichi. “Sporting performances and the volatility of listed football clubs.” International Journal of Sport Finance 6.4 (2011): 283.

[3] Benkraiem, Ramzi, Waël Louhichi, and Pierre Marques. “Market reaction to sporting results: The case of European listed football clubs.” Management Decision 47.1 (2009): 100-109.

[4] Bernile, Gennaro, and Evgeny Lyandres. “Understanding investor sentiment: The case of soccer.” Financial Management 40.2 (2011): 357-380.

[5] Bradley, Ralph Allan, and Milton E. Terry. “Rank analysis of incomplete block designs: I. The method of paired comparisons.” Biometrika 39.3/4 (1952): 324-345.

[6] Bunn, Christopher, et al. “Shirt sponsorship by gambling companies in the English and Scottish Premier Leagues: global reach and public health concerns.” Soccer & Society 20.6 (2019): 824-835.

[7] Peel, David & Cain, Michael, and Law, David. “The Favourite-Longshot Bias and Market Efficiency in UK Football Betting.” Scottish Journal of Political Economy. vol 47. 25-36. 10.1111/1467-9485.00151. (2000).

[8] Constantinou, A. C. and Fenton, Norman E.. “Profiting from arbitrage and odds biases of the European football gambling market.” The Journal of Gambling Business and Economics, Vol. 7, 2: 41-70. (2013).

[9] Dixon, Mark J., and Stuart G. Coles. “Modelling Association Football Scores and Inefficiencies in the Football Betting Market.” Journal of the Royal Statistical Society. Series C (Applied Statistics), vol. 46, no. 2, 1997, pp. 265–280. JSTOR, www.jstor.org/stable/2986290. Accessed 28 Mar. 2021.

[10] Easterbrook, Greg. “Broncos, Seahawks are battle-tested.” espn.com December 3. (2013).

[11] Franck, Egon, Erwin Verbeek, and Stephan Nüesch. “Inter-market arbitrage in betting.” Economica 80.318 (2013): 300-325.

[12] Gimet, Celine, and Sandra Montchaud. “What drives european football clubs’ stock returns and volatility?.” International Journal of the Economics of Business 23.3 (2016): 351-390.

[13] Alicja Gosiewska, Mateusz Bakala, Katarzyna Woznica, Maciej Zwolinski, and Przemyslaw Biecek. “EPP: interpretable score of model predictive power.“ arXiv:1908.09213 (2019).

[14] Grant, Andrew, et al. “New entry, strategic diversity and efficiency in soccer betting markets: the creation and suppression of arbitrage opportunities.” The European Journal of Finance 24.18 (2018): 1799-1816.

[15] Hess, J.A., “As Cristiano Ronaldo joins Juventus the team sells $60 million worth of his jerseys in 24 hours.” cnbc.com July 19. (2018).

[16] van Huet, Marc “Market Inefficiency in Soccer Betting Markets“ (2017).

[17] Jan Lasek, Zoltán Szlávik and Sandjai Bhulai. “The predictive power of ranking systems in association football.” International Journal of Applied Pattern Recognition (IJAPR), Vol. 1, No. 1 (2013).

[18] Malz, Allan M. “A simple and reliable way to compute option-based risk-neutral distributions.” FRB of New York Staff Report 677 (2014).

[19] Neal, Radford M. “MCMC using Hamiltonian dynamics.” Handbook of Markov Chain Monte Carlo 2.11 (2011): 2.

[20] Palomino, Frederic, Chendi Zhang, and Luc Renneboog. “Stock price reactions to short-lived public information: the case of betting odds.” ECGI-Finance Working Paper 81 (2005).

[21] Shimko, David (1993): “Bounds of Probability”, Risk, Vol. 6 (4) (April), pp. 33-37.

[22] Stan Development Team. “Stan Modeling Language Users Guide and Reference Manual.” https://mc-stan.org 2.19.3 (2021).

[23] Vlastakis, Nikolaos, George Dotsis, and Raphael N. Markellos. “How efficient is the European football betting market? Evidence from arbitrage and trading strategies.” Journal of Forecasting 28.5 (2009): 426-444.

[24] Winkelmann, David, Christian Deutscher, and Marius Ötting. “Bookmakers’ mispricing of the disappeared home advantage in the German Bundesliga after the COVID-19 break.” Applied Economics (2021): 1-11.

[25] Zuccolotto, Paola., Carpita, Maurizio., Sandri, Marco., and Simonetto, Anna. “Discovering the Drivers of Football Match Outcomes with Data Mining.” Quality Technology and Quantitative Management. 12. 537-553. 10.1080/16843703.2015.11673436. (2015).

A Supplementary Data Sources

Source	Description
Football-Data	Historical match outcomes across UK football divisions in 2005-2021 Closing betting odds for each match for up to 11 bookkeepers
EPL Team Payroll Tracker	Premier League player-level and club-level salary data
University of Glasgow	Shirt sponsorship by gambling companies in Premier League
Dimensions	Number of publications on sports betting arbitrage over time
Yahoo Finance API	Historical stock prices
Bloomberg	Historical option prices