Winners Restricted? A Comment on Kaunitz, Zhong and Kreiner

Elliot Christophers

Abstract

Kaunitz et al. (2017) employ a basic sports-betting strategy to generate positive returns, only to find their betting activity restricted by bookmakers. We evaluate their subsequent claim that it was their success which garnered the bookmakers' attention. We propose, and ultimately accept, an alternate hypothesis, namely, that it was instead the authors' insistence on betting only with the most generous bookmaker, in each individual case, which alerted the bookmakers and drew their restrictive response. This finding provides valuable advise for prospective bettors, as exploiting bookmakers only when they offer exposed prices is both more readily accomplished than being profitable, while also being more detrimental.

1 Introduction

In their 2017 paper "Beating the Bookies with their Own Numbers - and how the Online Sports Betting Market is Rigged", Kaunitz, Zhong and Kreiner (KZK) design a betting model which uses a wisdom-of-the-crowds effect to harness the information provided in the collection of market prices and bet profitably against the best price. After fitting their model, their backtests and eventual paper trades proved lucrative, prompting them to apply the model in practice with actual bookmakers. They placed a total of 265 bets, winning 125 of these, making a net profit of 957.5 units, from 50-unit stakes. This encouraging start was interrupted by their accounts with four separate bookmakers being restricted. KZK concluded that they had been too successful, since bookmakers are known to limit winning customers, and hence terminated their experiment. In this paper, we evaluate whether their betting data is sufficient evidence to support this claim, while proposing an alternate hypothesis, whereby KZK were instead restricted because they systematically bet only with the bookmaker with the longest price. We conclude that this second hypothesis is much more likely, offering guidance for bettors in that this conspicuous habit (which is easy to drop) is more harmful than being a winning customer.

2 The KZK Model

For a unit stake placed at odds o, if the bet wins, the profit will be $o-1$, while if it loses the entire stake is foregone. If the bet outcome has a probability p of occurring, the expected value for the bet is then $EV = p(o-1)+(1-p)(-1) = po-1$. For a fair bet, $EV = 0 = po-1 \implies p = \frac{1}{o}$. In general, then, the inverse of odds can be considered a noisy probability estimate for the underlying event.

The KZK betting model is what is known as, in some sections of the sports-betting literature, although they did not use the term, a quasi-arbitrage model (for discussion, see, for instance, Paton and Williams (2005), Smith et al. (2009) and Cortis (2016)). Essentially, the probability estimate p is constructed by considering some aggregate across the cross-section of odds offered by the various bookmakers. This probability is then compared with the longest (most profitable) odds in the cohort. It is therefore quasi-arbitrage because information is contained in the spread of market prices, which, if inefficient, can provide profitable trading opportunities.

If we let the different matches be indexed i and the outcomes be indexed j, then, for a certain match-outcome pair (I, J), we have that $O_{ij} = (o_{hij} \mid (i, j) = (I, J))$ is the tuple of odds offered by the different bookmakers, indexed h, on this match and outcome. We then have the longest odds $l_{ij} = \max(O_{ij})$ and the mean odds $\bar{o}_{ij} = \text{mean}(O_{ij})$. KZK let their quasi-arbitrage probability estimate for the j-th outcome of the i-th match be calculated $p_{ij} = \frac{1}{\bar{o}_{ij}} - c$ where c is a constant. Over their data, KZK found $c = 0.05$ to be the best fit. The expected value to betting at the longest odds l_{ij} is then

$$EV_{ij} = p_{ij} l_{ij} - 1 = \left(\frac{1}{\bar{o}_{ij}} - c \right) l_{ij} - 1 \tag{1}$$

where they would then bet on match I if

$$\begin{aligned} &EV_{IJ} > 0 \implies l_{IJ} > \frac{\bar{o}_{IJ}}{1 - c\bar{o}_{IJ}} \\ &\text{where} \\ &J = \arg\max_j (EV_{Ij}) \end{aligned} \tag{2}$$

Applying this betting strategy in practice, with actual bets with real bookmakers, KZK made a profit of 957.5 units from 265 bets with 50-unit stakes, winning 47,2% of their wagers. Despite these encouraging returns, the authors cut their experiment short due to being restricted by at least four bookmakers, where their maximum-allowed stakes were decreased to below the 50-unit equivalent. Indeed, KZK concludes, “Our study sets a precedent of the discriminatory practices against successful bettors in the online sports gambling industry: the online football market is rigged because bookmakers discriminate against successful clients.” Their sentiment is unequivocal: their betting was profitable, therefore, they were banned.

However, the relatively small sample size of 265 bets – especially considering that these wagers were spread across anywhere from four to thirty-two bookmakers – means that it is not necessarily certain that it was KZK’s profits which drew the bookmakers’ attention. Bookmakers do not want to make type-one errors, restricting what are fundamentally unprofitable customers, who have simply experienced a streak of good fortune. They will require significant evidence to ensure that the account being restricted is indeed a financial threat. We seek to evaluate whether the KZK betting history is indeed sufficient, or whether an alternative explanation is more convincing.

3 Evaluating the KZK Hypothesis

We thus want to evaluate the null hypothesis:

> H_0: A profit of 957.5 units from winning 47.2% of 265 50-unit bets across four to thirty-two bookmakers is not sufficient evidence for a bettor to be restricted by four or more bookmakers. (3)

To evaluate this hypothesis, we would first need to develop a framework within which bookmakers can make decisions about their customers’ potential as sustainable winners. Tryfos et al. (1984) developed a statistical test for assessing the null hypothesis that an American football betting model is unprofitable. An invocation of the central limit theorem, Tryfos et al. essentially produce a z-test tailored for American football betting. For our purposes, where the sample size, in this case, faced by each bookmaker will be very small, a one-sample t-test (Student, 1908) seems more appropriate.

Consider a betting history where n bets, indexed k, are made at odds o_k. Let $\boldsymbol{I}(k)$ be an indicator function

$$\boldsymbol{I}(k) = \begin{cases} 1 & \text{if the } k\text{-th bet is won} \\ 0 & \text{otherwise} \end{cases} \tag{4}$$

then $R = (\boldsymbol{I}(k)o_k - 1 \mid k = 1, \ldots, n)$ is the n-tuple of returns to unit stakes, for the betting history. We can then calculate the mean $\bar{r} = \frac{1}{n}\sum_{r_k \in R} r_k$ and variance $s^2 = \frac{1}{n-1}\sum_{r_k \in R}(r_k - \bar{r})^2$ of returns, and evaluate the null hypothesis $H_0 : r = 0$ using the aforementioned t-test, where r is the betting model's true rate of return to unit stakes, estimated by $\bar{r}$, with s^2 the estimator for the variance of r. If a bettor has placed all of these n bets with a certain bookmaker, the bookmaker could then evaluate the one-sided alternative hypothesis $H_1 : r > 0$ based on $\bar{r}$ and s^2, and, if the resultant test-statistic is sufficiently high, restrict the bettor.

We unfortunately know neither the R nor even the n seen by any of the bookmakers which laid bets for KZK. Instead, we will have to use Monte Carlo simulation to produce different possible betting histories which match the KZK betting data. Although we do not have access to the exact dataset analysed by KZK, we assume that the distribution of odds when betting with the same model on a different but similar dataset will approximate the distribution of o_k generated by their model in (2). We thus access the football-data.co.uk closing-odds dataset, 5 seasons' worth of soccer games from 22 of the biggest leagues across Europe, spanning the 2019/20-2023/24 seasons. For each of the 37,969 games, the data contains closing odds from as many as six bookmakers, for each outcome.

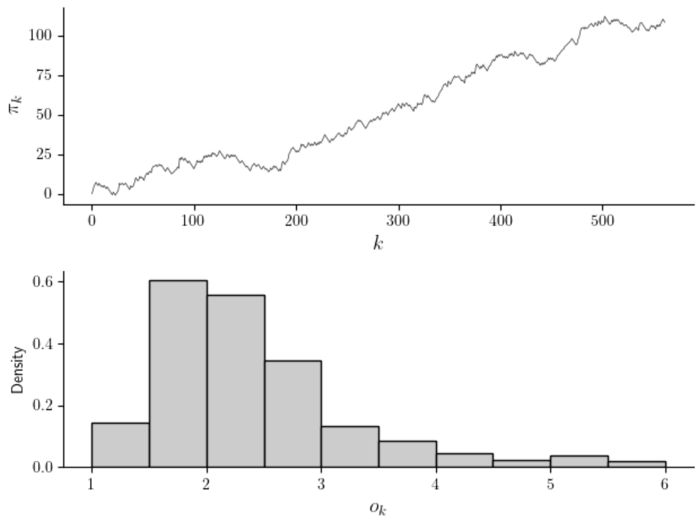

Figure 1: Upper panel: using the KZK model defined by (2), a profit of 108.29 units is made from 561 bets over our dataset. Lower panel: the distribution of o_k, the odds bet at in the upper-panel trajectory.

Without inspecting the data, we set $c = 0.05$ (as had KZK) and apply the KZK betting model in (2) to this data. Letting π_k be the profit to unit stakes after k bets, made at odds o_k, the upper panel in Figure 1 shows the profit trajectory, while the lower panel shows the distribution of the o_k. A mere 561 games, 1.48% of the available 37,969, are selected for betting. Across these, a profit from unit stakes of 108.29 units is achieved, while a p-value from the t-test described above of 0.000173 is obtained, demonstrating the robustness of the KZK model. If these 561 bets were all placed with the same bookmaker, according to our framework, they would restrict this bettor with any significance level $\alpha \geq 0.000173$.

It is unclear exactly what significance level a bookmaker would set, not wanting to accidentally ban a fundamentally losing customer. A washingtonpost.com (2022) article suggested that as many as 10% of customers can be affected by account restrictions. A discussion on the (admittedly biased) justiceforpunters.org blog relating to a survey conducted by Horseracing

Bettors Forum (2016) indicated that roughly 4% of inexperienced bettors and upwards of 10% of experienced bettors are restricted. The CEO of a large UK bookmaker disclosed that 3% of their customers are restricted (sbcnews.co.uk, 2018). We can then imagine that the true proportion of successful bettors lies in the 1-10% range, and that this, therefore, would also be a reasonable range for bookmakers to set their significance level α. Therefore, clearly, we can confirm that the KZK returns are significant, and the model is indeed profit-generating. However, the above p-value is facing only one bookmaker, whereas the KZK results were distributed across at least four bookmakers, and as many as thirty-two.

To test the null hypothesis in (3), we therefore perform a Monte Carlo simulation. As mentioned, we do not know what odds o_k KZK bet at, however, we can assume that their distribution was approximately the same as the lower panel in Figure 1. Considering the definition of R above, we can see that only the odds for won bets contribute to the profit. Therefore, we want to generate a tuple of 125 odds (the number of bets won by KZK) which, together with 140 lost bets, would give a profit of 957.5 units from 50-unit stakes, while having approximately the same distribution as the o_k from the KZK run across our dataset. Let $O' = (o'_k \mid k = 1, \dots, 125)$ be this generated tuple of 125 odds, and $O = (o_k \mid k = 1, \dots, 561)$ be the tuple of odds o_k from Figure 1. We construct O' by minimising the earth mover's distance (Rubner et al., 1998), also (for probability distributions) known as the Wasserstein distance (Vaserstein, 1969), between O' and O, subject to O' generating the observed profit of 957.5 units, as described. Across the simulation, we want O' to vary in different iterations. Therefore, we provide different initial guesses in each case, O_0 being a 125-tuple of randomly selected elements from O. Figure 2 shows nine iterations of O'. Notice the consistent similarity to the lower panel in Figure 1, despite the variation and individual features in each.

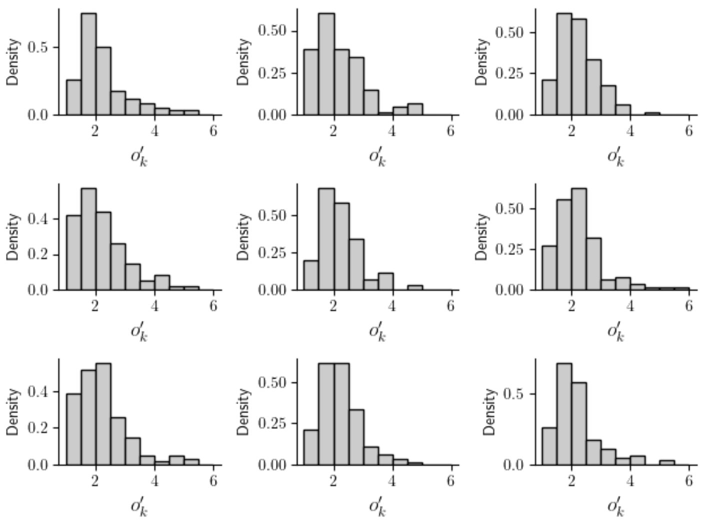

Figure 2: Nine iterations of O', 125-tuples of odds which give the observed KZK profit while minimising the earth mover's distance to O, the tuple of odds from the KZK model on our dataset.

There remains uncertainty regarding four important features of the KZK betting results: how many bookmakers did they bet with and how many bets did they place with each, what significance levels do the bookmakers have, and how many bets will a bettor have to place before being analysed for profitability. In each iteration i of the simulation, we handle these questions stochastically. The KZK dataset spanned thirty-two bookmakers, therefore, each iteration sees a random integer m_i in the range $[4, 32]$ drawn uniformly to represent the number of bookmakers, indexed h, bet with. We set 4 as the lower bound here because we know for certain that this is the number of bookmakers with which KZK placed bets. Supplementary Figure 1 in KZK shows that paper trades were placed with four further bookmakers, while a footnote indicates that real bets were probably placed with one further. However, as a lower bound here, we take 4, due to the certainty, while this is also the minimum number of bookmakers needed to restrict KZK for an iteration to give a success, meaning an even lower bound would be foolish. Different bets are considered independent, giving a m_i-tuple of probabilities (p_{hi}) which sum to one, also drawn uniformly, representing the probability of each bet being placed with the corresponding bookmaker h. Similarly, (α_{hi}) is the m_i-tuple of significance levels for the m_i bookmakers, each drawn uniformly from the range $[0.01, 0.1]$ (see the discussion above).

Regarding the number of bets before a bookmaker might consider a bettor as profitable, purely statistically, it makes little sense to apply a t-test after only a few bets. Similarly, the survey mentioned above (Horseracing Bettors Forum, 2016) gave 100, 50 and 20 (all, presumably, referring to a range of bets, rather than an exact threshold) as common numbers of bets preceding bookmaker intervention. Perhaps inconsistently, given these facts, we set three thresholds, $\tau = \{5, 15, 30\}$, as the minimum number of bets necessary. When the m_i is large, and the p_{hi} are relatively equal, it is unlikely that any individual bookmaker will lay a dominating proportion of the 265 bets, let alone four bookmakers. We perform simulations with these three values for τ, being interested in the results despite recognising that $\tau = 5$ is not particularly realistic.

We also vary the nature of these other unknowns. In certain simulations, we will draw m_i as described above, while in others it will be fixed at four, the smallest number of bookmakers which we know bets were placed with, which therefore intuitively maximises the probability of rejecting the H_0 in (3). Similarly, certain simulations will see the $p_{hi} = \frac{1}{m_i} \; \forall \; h$, that is, each bookmaker is equally likely to have laid a given bet, while some simulations will have $\alpha_{hi} = 0.05 \; \forall \; h$, a standard significance level. We thus create one simulation for each of the combinations of treatments of the unknowns, all in all giving twenty-four simulations ($2 \times 2 \times 2 \times 3 = 24$, for the three binary choices of random or determined assignment, and then the three thresholds). Let $\tilde{m}, \tilde{p}$ and $\tilde{\alpha}$ be binary variables for whether these assignments are random or determined, with 1 being random and 0 determined. Note that $\tilde{p} = 0$ will still give random p_{hi} in the sense that $\frac{1}{m_i}$ will vary if $\tilde{m} = 1$.

In each simulation, 1,000 iterations are performed, drawing O_i' and allocating the 125 odds and 140 losses to the different bookmakers as established by the simulation parameterisation $\theta = (\tilde{m}, \tilde{p}, \tilde{\alpha}, \tau)$. For each bookmaker, if the number m_{hi} of bets faced is $m_{hi} \geq \tau$, then the t-test described above is applied and if the resulting p-value is less than α_{hi}, the bettor is restricted. The number of restrictions across the m_i bookmakers are counted, and if this number is greater than or equal to four, the iteration is a success, otherwise it is a failure. The overarching p-value p_θ corresponding to the overall hypothesis H_0 from (3) is the proportion of failures from the 1,000 iterations. Table 1 shows the results from this simulation.

$\tilde{m}$	$\tilde{p}$	$\tilde{\alpha}$	τ	p_θ
0	0	0	5	0.529
			15	0.543
			30	0.528
		1	5	0.589
			15	0.571
			30	0.618
	1	0	5	0.796
			15	0.821
			30	0.885
		1	5	0.810
			15	0.832
			30	0.891
1	0	0	5	0.245
			15	0.693
			30	0.906
		1	5	0.225
			15	0.687
			30	0.898
	1	0	5	0.334
			15	0.727
			30	0.935
		1	5	0.293
			15	0.682
			30	0.951

Table 1: P-value p_θ for the null hypothesis in (3), for twenty-four different simulations, each parameterised by $\theta = (\tilde{m}, \tilde{p}, \tilde{\alpha}, \tau)$.

Despite there being twenty-four different simulations, each with a different combination of parameters in θ, the lowest p-value p_θ is 0.225, not low enough to reject the null hypothesis in any case. Table 2 shows the mean p-value across each parameter value. It is immediately evident that the value of τ is especially critical. As discussed above, $\tau = 5$ is unreasonable both from a statistical perspective (applying a t-test with a sample-size of 5) and from a practical perspective (bookmakers will not ban a winning customer after only five bets). In Table 1, the four lowest p-values, as well as the sixth, all have $\tau = 5$. The smallest value of p_θ with $\tau \neq 5$ is 0.528, clearly insufficient to reject H_0. The remaining results are not particularly interesting. In the majority of simulations, four or more bookmakers reject KZK in 20-40% of iterations. Four or more bookmakers are less likely to restrict the bettor when the the number of bookmakers m_i is fixed than randomly drawn, while the reverse is true for the probability p_{hi} of a bookmaker having laid each bet. Ultimately, H_0 cannot be rejected. The alternative, however, that KZK were restricted because of their betting success, is not entirely unlikely, being a 20-40% probability event. In the next section we consider a second hypothesis, that KZK were restricted due to exclusively betting with the longest bookmaker.

		$\bar{p}_\theta$
$\tilde{m}$	0	0.701
	1	0.631
$\tilde{p}$	0	0.586
	1	0.746
$\tilde{\alpha}$	0	0.662
	1	0.671
τ	5	0.478
	15	0.695
	30	0.827

Table 2: $\bar{p}_\theta$ is the mean p-value across all simulations with the given parameter value.

4 An Alternate Hypothesis

The KZK betting strategy relies upon betting only with the bookmaker which offers the most generous odds, in each individual case. The ability to execute these trades, then, requires the knowledge of which bookmaker has the longest odds. In the same way that KZK know this, the bookmakers will too, and can therefore recognise that KZK only bet with them when they hold the most vulnerable position in the market. As a bookmaker, clearly, if a customer is winning, and only bets when your prices are the most favourable, this would be a much stronger signal that the customer is dangerous, compared to them merely being winning. The distinction is that a losing customer can be winning even after a reasonable number of bets, but a recreational bettor cannot only select bets where their bookmaker has the longest odds purely by chance.

Suppose that there exist μ bookmakers, where an active bettor tracks the prices of $m \leq \mu$ bookmakers, while the h-th bookmaker tracks $m_h \leq \mu$ bookmakers (themselves included). Let $H = \{h \mid h = 1, \ldots, \mu\}$ be the set of indices for all bookmakers. Then, let $T = \{h_j \mid j = 1, \ldots, m; h_j \in H \; \forall \; j\}$ be the set of indices representing the bookmakers tracked by the bettor, ordered by their global indices h, and let $T_h = \{k_{hj} \mid j = 1, \ldots, m_h; k_{hj} \in H \; \forall \; j\}$ be the set of indices tracked by the h-th bookmaker, similarly ordered. For a given outcome, all μ bookmakers offer odds o_h, with duplicate prices possible. Let Q be the m-tuple $Q = (o_j \mid j \in T)$ and Q_h the m_h-tuple $Q_h = (o_j \mid j \in T_h)$ be the tuples with the odds offered by those bookmakers tracked by the bettor and bookmakers, respectively.

Let R be the set where the j-th element of R gives the rank of o_{h_j} in Q, and R_h be the set where the j-th element of R_h gives the rank of $o_{k_{h_j}}$ in Q_h. In the case of ties, ranks are assigned as the higher value, for instance, if $m = 3$, with $o_{h_2} < o_{h_1} = o_{h_3}$, then $R = (3, 1, 3)$. Then, for a given profitable betting opportunity, the gambler will bet with any of the bookmakers h_j which satisfy $R_j = m$, where R_j denotes the j-th element of R. If a bettor places several bets with bookmaker $h_j \in T$, which maps to $h \in H$ which maps to $k_{hj'} \in T_h$, the bookmaker

h will become suspicious if $m_h - R_{h,j'}$ is consistently close to zero, where $R_{h,j'}$ denotes the j'-th element of R_h. Indeed, if the bettor places n bets with bookmaker h, then we can denote $\Psi_h = (R_{hl,j'} \mid l = 1, \ldots, n)$ as the n-tuple of ranks, from the bookmaker's perspective, of their odds. We can then imagine that the bookmaker might want to evaluate the hypothesis

$$H_0 : \psi_h = (1 - \Delta_h)m_h \tag{5}$$

where ψ_h is some aggregate representation of Ψ_h, and $\Delta_h \in [0, 1]$ is some decimal defining a percentile $1 - \Delta_h$ of the number of bookmakers m_h which the h-th bookmaker's aggregate rank ψ_h is tested against. For instance, the bookmaker might want to test the hypothesis that the bettor bets with them when their odds are in the 80th percentile, in this case setting $\Delta_h = 0.2$. It seems reasonable to use a non-parametric test to evaluate (5), where a one-sided Wilcoxon signed-rank test (Wilcoxon, 1992) is the most suitable.

The hypothesis in (5) is what the individual bookmakers will test to evaluate whether the bettor bets only on relatively (compared to the market) long odds. Similar to how Monte Carlo simulation was used to assess H_0 from (3) above, we will again use simulation to evaluate

> H_0: When a bettor considers the odds offered by $m \leq \mu$ bookmakers, who each consider $m_h - 1 \leq \mu$ of their peers, if the bettor places 265 bets with whomever of the m bookmakers offer the longest odds in each case, this – coupled with the bettor performing reasonably well returns-wise – is not sufficient evidence for them to be restricted by four or more bookmakers, all making decisions according to (5). (6)

For KZK, we know that $m = 32$. However, if they only bet with some $M < 32$ different bookmakers across their 265 bets, then it is as if $m = M$. In each iteration i, therefore, we draw $m_i \in [4, 32]$, where m_i is an integer. Given this, we similarly draw the total number of bookmakers $\mu_i \in [m_i, 100]$, with a high upper bound given the manifold bookmakers which do exist, giving $H_i = \{h \mid h = 1, \ldots, \mu_i\}$. Hence, we select $m_{hi} \in [\min(m_i, 10), \mu_i]$, the bookmaker counts for the μ_i bookmakers indexed h (10 seems a reasonable lower bound for the m_{hi} when $m_i > 10$), and create $T_{hi} = \{k_{hi} \mid k_{hi} \in H_i \ \forall \ h; \exists \ k_{hi} = h_i\}$ by randomly selecting $m_{hi} - 1$ indices $h \in H_i$, in each case, ensuring that $h \in T_{hi}$ for all $h = 1, \ldots, \mu_i$. As in the above simulation, we generate two μ_i-tuples, one of probabilities $P_i = (p_{hi})$ which sum to one for assigning odds ranks, and one for significance levels $(\alpha_{hi} \mid \alpha_{hi} \sim U(0.01, 0.1) \ \forall \ h)$ used in assessing (5). Also for this latter purpose, we create $((1 - \Delta_{hi}) \mid \Delta_{hi} \sim U(0.1, 0.5) \ \forall \ h)$, a μ_i-tuple of percentiles which the bookmakers compare the bettor's bet selection to. Finally, we also generate the 125-tuple O'_i, as was done above.

Given these objects, an iteration is completed as follows. For each of the 265 bets, we draw a μ_i-tuple $P'_i = (p'_{hi} \mid p'_{hi} \sim U(0, 1) \ \forall \ h)$. The rank of odds across the μ_i bookmakers is then the rank of the μ_i-tuple resulting from the component-wise product of P_i and P'_i. The bettor then considers the bookmakers in T_i and creates Q_i and R_i, placing a bet with any bookmaker $h_j \in T_i$ with $R_{i,j} = m_i$. This bookmaker maps to $h \in H_i$, which maps to $k_{h,j'} \in T_{hi}$. The h-th bookmaker notes down $R_{hi,j'}$, as well as the return of the bet, which is either some $o_i \in O'_i$ or else -1 (selected randomly without replacement). After these 265 bets, each bookmaker h who has laid at least one bet constructs Ψ_{hi} and evaluates (5) with respect to α_{hi}, $(1 - \Delta_{hi})$, and the observed $R_{hi,j'}$. If the one-sided p-value from the Wilcoxon signed-rank test is less than α_{hi}, and the bettor's mean return is positive, the bettor is restricted. The number of restrictions across the μ_i bookmakers are counted, and if this number is greater than or equal to four, the iteration is a success, otherwise it is a failure. The p-value p_θ corresponding to the overall hypothesis H_0 from (6) is the proportion of failures across all iterations. Also as above, we define binary simulation-wide parameters $\theta = (\tilde{m}, \tilde{m_h}, \tilde{p}, \tilde{\alpha}, \tilde{\Delta})$, which determine whether the various objects above are randomly generated. If $\tilde{m} = 0$, then $m_i = 4 \; \forall \; i$. If $\tilde{m_h} = 0$, then $m_{hi} = \min(m_i, 10) \; \forall \; (h, i)$. If $\tilde{p} = 0$, then $P = \left(\frac{1}{\mu_i}\right) \; \forall \; i$. If $\tilde{\alpha} = 0$, then $\alpha_{hi} = 0.05 \; \forall \; (h, i)$. Finally, If $\tilde{\Delta} = 0$, then $\Delta_{hi} = 0.25 \; \forall \; (h, i)$. This parameterisation gives thirty-two simulations. With 1,000 iterations in each, as above, the results are displayed in Tables 3 and 4.

$\tilde{m}$	$\tilde{m_h}$	$\tilde{p}$	$\tilde{\alpha}$	$\tilde{\Delta}$	p_θ
0	0	0	0	0	0.004
				1	0.452
			1	0	0.004
				1	0.413
		1	0	0	0.878
				1	0.922
			1	0	0.876
				1	0.910
	1	0	0	0	0.082
				1	0.599
			1	0	0.101
				1	0.602
		1	0	0	0.919
				1	0.916
			1	0	0.911
				1	0.920
1	0	0	0	0	0.000
				1	0.024
			1	0	0.001
				1	0.020
		1	0	0	0.104
				1	0.128
			1	0	0.119
				1	0.117
	1	0	0	0	0.001
				1	0.031
			1	0	0.002
				1	0.025
		1	0	0	0.116
				1	0.125
			1	0	0.110
				1	0.121

Table 3: P-value p_θ for the null hypothesis in (6), for thirty-two different simulations, each parameterised by $\theta = (\tilde{m}, \tilde{m_h}, \tilde{p}, \tilde{\alpha}, \tilde{\Delta})$.

We can immediately see that, in general, the p-values p_θ are lower in Table 3 than in Table 1, suggesting that the null hypothesis in (6) is more likely to be rejected than that in (3). Indeed, whereas none of the thirty-two combinations of parameters θ under the KZK hypothesis led to a rejection of that null for any reasonable significance level, ten of thirty-two parameter combinations reject the alternate hypothesis discussed in this section at $\alpha = 0.05$ and six do so for $\alpha = 0.01$. Indeed, each of the ten possible parameter values (see Table 4) contribute to a simulation with p_θ less than the smallest value observed in Table 1, with 0.104 being the largest such value, the smallest value of p_θ when $\tilde{p} = 1$.

Considering Table 4, interestingly, it is easier for bookmakers to reject the null hypothesis in (5) when testing if a bettor bets only when the their odds are above the 75-th percentile, than for a random percentile drawn between 50 and 90. This is encouraging, as high power against this reasonably high percentile, compared to the alternative where the majority of bookmakers consider $(1 - \Delta_{hi}) < 0.75$, suggests that bettors who exploit only favourable pricing are, in fact, generally restricted. The two most drastic splits occur in $\tilde{m}$ and $\tilde{p}$. In the case of the former, when m_i is constrained to equal 4, the mean p-value p_θ equals 0.594, considerably larger than the 0.065 observed when m_i can be upwards of 32. This is encouraging, as it is unlikely that KZK only bet with four bookmakers: we merely know that they bet with at least four, making this the natural lower-bound to test in the $\tilde{m} = 0$ case. Conversely, as intimated above, $\hat{p}_\theta$ is substantially larger for $\tilde{p} = 1$, with at least four bookmakers much more likely to restrict the KZK bettor when all bookmakers are equally likely to offer the longest odds on each game. This is concerning, as the alternative, where certain bookmakers are systematically more likely to be the longest, a feature which $\tilde{p} = 1$ captures, does seem more faithful to reality: Grant et al. (2018) and Deschamps and Gergaud (2007), for instance, provide evidence in favour of this theory.

Both here and above, the parameters $\tilde{m}, \tilde{p}$ and $\tilde{\alpha}$ have the same purpose. We find the same split in $\tilde{p}$ in Table 2, with $\tilde{p} = 0$ giving lower p-values,

		$\bar{p}_\theta$
$\tilde{m}$	0	0.594
	1	0.065
$\tilde{m}_h$	0	0.311
	1	0.349
$\tilde{p}$	0	0.148
	1	0.512
$\tilde{\alpha}$	0	0.331
	1	0.328
$\tilde{\Delta}$	0	0.264
	1	0.395

Table 4: $\bar{p}_\theta$ is the mean p-value across all simulations with the given parameter value.

$\tilde{m}$	$\tilde{p}$	$\tilde{\alpha}$	(3)	(6)
0	0	0	0.529	0.267
		1	0.589	0.257
	1	0	0.821	0.918
		1	0.832	0.911
1	0	0	0.693	0.013
		1	0.687	0.011
	1	0	0.727	0.121
		1	0.682	0.118

Table 5: For those parameters in θ which exist in (3) and (6), we compare the median p_θ across the simulations with the given parameter values.

on average. Table 5, then, compares the values of p_θ for the combinations of values for these three parameters, when used in evaluating (3) and (6), respectively. Each combination occurs in several simulations, and since these depend on the other parameters τ (which had especially drastic splits in Table 2 above), as well as $\tilde{m_h}$ and $\tilde{\Delta}$, respectively, we consider the median value of p_θ, rather than the mean, which was used in Tables 1 and 3, when the comparisons were intra-table, rather than inter-table. The only two scenarios where p_θ is lower with the KZK hypothesis is for $(\tilde{m}, \tilde{p}) = (0, 1)$. As discussed above, $\tilde{m} = 0$ is not a particularly realistic scenario (probably less likely than $\tilde{p} = 0$, considering the discussion above of KZK most likely betting with more than only four bookmakers), which suggests that we need not necessarily be distressed by $\tilde{p} = 1$ giving higher p-values, especially considering how, in these two cases when the KZK p-values are lower than for the alternate hypothesis, they are both above 0.8, not remotely near rejecting the KZK hypothesis. Indeed, we are concerned not with, necessarily, rejecting either of the hypotheses in (3) or (6); rather, we are concerned with evaluating which is more likely to be rejected. The evidence in this section, we believe, is sufficiently compelling to demonstrate that it is more likely that KZK were restricted due to betting only with the longest bookmaker than because they were profitable bettors.

5 Conclusion

After a brief period of successful betting on soccer, with profits very much aligned with expectations formed from backtests and paper trades, Kaunitz et al. (2017) found themselves restricted from further betting with four separate bookmakers. Their conclusion was that the bookmakers were simply discriminating against a successful bettor (a discretionary practice not at all unheard of), and hence argued that this power is unfair, anti-competitive, and the like. However, given the ephemeral nature of their actual betting with the bookmakers, we decided to scrutinise this hypothesis while proposing another, namely, that it was instead the authors' behaviour of only betting with a certain bookmaker if they offered the most generous price in the market which prompted the restrictions. After defining null hypotheses to the extent of each of these positions, we set up a Monte Carlo simulation for each, producing a set of p-values against these hypotheses under various conditions. Running these simulations against the betting data described by the authors, we determine that it is much more likely that they were restricted because they exclusively transacted with the most generous bookmaker than simply because they were profitable, which ultimately can occur in the short run even for what are fundamentally losing bettors.

6 References

Cortis, Dominic. “Betting Markets: Defining odds restrictions, exploring market inefficiencies and measuring bookmaker solvency”. 2016. U of Leicester, PhD dissertation.

Deschamps, Bruno, and Olivier Gergaud. “Efficiency in betting markets: evidence from English football”. *The Journal of Prediction Markets*, vol. 1, no. 1, 2007, pp. 61–73.

football-data.co.uk. “Data Files: All Countries”, www.football-data.co.uk/downloadm.php. Accessed 15 Jan. 2024.

Grant, Andrew R., et al. “New entry, strategic diversity and efficiency in soccer betting markets: the creation and suppression of arbitrage opportunities”. *The European Journal of Finance*, vol. 24, 2018, pp. 1799–816. https://doi.org/10.1080/1351847X.2018.1443148.

Horseracing Bettors Forum. “Further details of HBF’s Account restriction/Closure Survey”, 2016, ukhbf.org/account-restrictionclosure-survey/further-details-of-hbfs-account-restrictionclosure-survey/. Accessed 9 Aug. 2024.

justiceforpunters.org. “How many people have restricted betting accounts?”, justiceforpunters.org/campaigning/how-many-people-have-restricted-betting-accounts/. Accessed 9 Aug. 2024.

Kaunitz, Lisandro, et al. “Beating the bookies with their own numbers-and how the online sports betting market is rigged”. *arXiv preprint arXiv:1710.02824*, 2017.

Paton, David, and Leighton Vaughan Williams. “Forecasting outcomes in spread betting markets: Can bettors use ‘quarbs’ to beat the book?” *Journal of Forecasting*, vol. 24, no. 2, 2005, pp. 139–54.

Rubner, Yossi, et al. “A metric for distributions with applications to image databases”. *Sixth international conference on computer vision (IEEE Cat. No. 98CH36271)*. IEEE, 1998, pp. 59–66.

sbcnews.co.uk. “Sky Bet lifts the lid on account restrictions at Parliamentary debate”, 2018, sbcnews.co.uk/sportsbook/2018/01/24/sky-bet-lifts-lid-account-restrictions-parliamentary-debate/. Accessed 9 Aug. 2024.

Smith, Michael A, et al. “Do bookmakers possess superior skills to bettors in predicting outcomes?” *Journal of Economic Behavior & Organization*, vol. 71, no. 2, 2009, pp. 539–49.

Student. “The probable error of a mean”. *Biometrika*, 1908, pp. 1–25.

Tryfos, Peter, et al. “The profitability of wagering on NFL games”. *Management Science*, vol. 30, no. 1, 1984, pp. 123–32.

Vaserstein, Leonid Nisonovich. “Markov processes over denumerable products of spaces, describing large systems of automata”. *Problemy Peredachi Informatsii*, vol. 5, no. 3, 1969, pp. 64–72.

washingtonpost.com. “Sportsbooks say you can win big. Then they try to limit winners.” 2022, www.washingtonpost.com/sports/2022/11/17/betting-limits-draft-kings-betmgm-caesars-circa/. Accessed 9 Aug. 2024.

Wilcoxon, Frank. “Individual comparisons by ranking methods”. *Breakthroughs in statistics: Methodology and distribution*, Springer, 1992, pp. 196–202.