Efficiency, Bias, and Decisions: Observations from a Sports Betting Exchange

Alexander Kan

University of California, Berkeley

May 15, 2020

latest version here

Abstract

We examine the efficiency of sports wagering markets in a betting exchange and find that they serve as good predictors of true outcomes, but do have a bias in which favorites are undervalued and longshots are overvalued. We consider work on the bias spanning behavioral and structural justifications for its existence, and focus on access to information as well as prospect theory in our analysis. The results from sports betting exchanges in this paper suggest that the existence of the bias is not due to information or transaction costs, implying that work involving sportsbook structure may not accurately reflect market behavior. Further, we show that the bias is not present in bets that were taken prior to the start of a sporting event but is prevalent in bets that take place after it begins. We conclude that more informed bets may be reacting suboptimally to information, and that individuals may be making irrational weighting decisions akin to results found in analysis of prospect theory.

Keywords: Market Efficiency, Sports Wagering, Favorite-Longshot Bias, Prospect Theory

Acknowledgements: Special thanks to Professor Raymond Hawkins, Todd Messer, as well as my family and friends for support, advice, input, and discussion that aided my efforts.

1 Introduction

The traditional hypothesis of efficient markets suggests that the market price for a security incorporates all potential information available, and thus it is not possible to predict what will happen better than the market. As such, in the long run one cannot consistently beat the market. Analysis of efficiency frequently revolves around showing that the values of a security follow a stochastic process, to suggest that the information added to the market is random. The sports markets have potential to be useful due to the fact that each wager has a true and observable value following the conclusion of an event. Whereas it is unclear how much a security should truly be worth, a sport bet either wins or loses, and following the event the result is known. In effect, sports bets are similar to binary options, and examining their behavior can give insight on other financial markets. As a result, sportsbook lines have been previously used to test efficiency. However, sportsbooks are controlled by casinos, who adjust lines at their will and can accept or reject any given bet based on business needs. Instead, the recent rise of betting exchanges is much more representative of a financial market as they allow customers to exchange freely with others as well as close or add to their positions at any time in any amount. Crucially, the use of an exchange allows bettors to choose either side of any given bet.

A sports betting exchange operates much like a financial exchange, and consists of two primary actors, 'backers' and 'layers'. Backers pay to take on a bet at certain odds, and layers would take the opposite end of that bet. For instance, suppose a layer is offering a \$10 bet at odds of 3 (European Style Odds) that team X will win its match against team Y. A backer, can then choose to take this bet for up to \$10. In the case that team X does win, the backer wins their bet, and the layer would pay out \$30 to the backer who would also recieve their originally wagered \$10 (Total return of \$40). If instead, X does not win, the layer gets the \$10 that the backer placed their bet with. The role of the exchange is simply to facilitate this transaction as well as to show both backers and layers the best bids and offers that exist in the market. For this service, exchanges charge a commission, in the case of Betfair, the world's largest betting exchange, this commission is up to 5% of the total profit, an edge that is significantly smaller than traditional sportsbooks.

As a result, sports exchanges can serve as a better avenue to test efficiency than sportsbooks since they more closely resemble a financial market and examining these markets can give more insight on analogous exchanges in financial products. The data used in this paper comes from Betfair, and contains a one-week cross section of all bets placed through the exchange over a variety of sports. It contains details on the odds at which a bet was traded, as well as the final outcome of the bet, either a win or a loss. It also has details for trades that occurred before a sporting event started, as well as in play betting which takes place during the event, a concept known as live-betting.

The aim of this paper is to determine whether or not these markets are efficient. More specifically, our test of efficiency, is really a test of whether or not a classifier based on market probabilities is calibrated. That is for an outcome, $Y \in \{0, 1\}$, and a market probability of an event X, for every $r \in [0, 1]$

$$P(Y = 1|X = r) = r \ . \tag{1}$$

If this condition is satisified, the probabilities implied by the odds at which events are trading, match the observed probability of an event occurring after the event concludes. This would mean that odds produced by the market reflected their true value and are efficient. Our hypothesis is that sports exchange markets will tend to be efficient and this paper will seek to discover why this hypothesis does or does not hold and try to provide rationale for potential deviations.

2 Literature Review

Markets on sports have become a focus for economists primarily due to the unique fact that the outcome of a sports event is known after its conclusion, and therefore, the true value of the wager can be found. As Sauer (1998) explains, betting markets are simple versions of financial markets that exhibit similar properties but are easier to examine. Sauer extensively analyzes horse racing markets and ultimately finds that these markets are mostly efficient, and that they effectively predict the probability of a horse winning a race. In fact, the general consensus in the literature is that markets are good predictors of true probabilities and as a result, this statement is often just assumed as fact in literature.

However, Sauer (1998) shows that there is a notable anomaly that exists, known as the favorite-longshot bias, in which the prices of favorites are undervalued, while longshots are overpriced. That is, for events that tend to be unlikely to happen, the market suggests that they would happen more frequently than they do in actuality, with the reverse being true for events that are likely to occur. This bias has become the focus of a lot of the research in the realm of market efficiency. Two major schools of thought on the root of the favorite-longshot bias have emerged. The first regarding risk preferences of bettors, and the second considering institutional forces.

Quandt (1986) explores the risk preferences of bettors and suggests that the fact that bettors are willing to make decisions that they know are negative in expectation implies that they must be risk seeking individuals. He then suggests that because these bettors are risk seeking, they should simply bet on whichever horse has the highest variance. However, in practice this doesn't occur, otherwise all but one horse would have zero bets placed on it. As such, Quandt suggests that it is necessary for some bias to exist in order to reach an equilibrium in the markets.

Thaler and Ziemba (1988) suggested a variety of behavioral reasons such a deviation from expectation may exist at the horse racing tracks they examined. They argue that there is more enjoyment that comes from betting on a longshot than a favorite, as winning on a longshot simply gives bettors a better story to tell than winning on a favorite. They also suggest that some bettors just make decisions on an irrational basis, based on something like the name of a horse. Finally, they suggest an effect similar to observations in Tversky and

Kahneman (1992), which finds that decision weights are not linear with true probabilities, and that individuals underweight high probability events and overweight low probability events. Further, Tversky and Fox (1995) expands on this by explaining that jumps in probability from an event being highly likely to becoming certain are more impactful than equivalent jumps from the event being likely to slightly more likely.

Alternatively, other research has founnd explanations using empirical models for the bias by examining institutional effects, such as the differing access to information the bettors have, as well as transaction costs and the response of sportsbooks to informed bettors. Shin (1992) writes of the existence of insiders in the markets, and concludes that bookmakers create the favorite-longshot bias intentionally to pass on the losses of informed bettors to those who are uninformed. Shin assumes that without the existence of insider trading, the market's probability of a horse winning would be identical to the true probability (i.e. the markets are efficient). He then conducts an optimization for the bookmaker profit, and finds that given that insiders do exist, it is most profitable for bookmakers to have prices that undervalue favorites and overvalue longshots. Sobel and Raines (2003) construct models of both risk preferences and information, and ultimately find that there is little variation in the bias over bets with different risk profiles, but that variation of information does in fact create a differing level of bias in deviations from expectation.

Meanwhile, Hurley and McDonough (1995) consider an experimental approach examining transaction costs, and how they impact the decisions that bettors make. They assert that without these costs, bettors could calculate the true probabilities, and that the costs attached to betting by the sportsbook create a deviation between the subjective and objective probabilities. They take this a step further and suggest that in the case where bettors are entirely uninformed they should bet with equal probability on each event, creating a situation in which they over-bet on longshots, and under-bet on favorites. Then, as the transaction costs inhibit access to information, higher transaction costs mean fewer informed bettors and therefore more of a bias. They further this analysis with two experiments that test behaviors of bettors in an environment with and without transaction costs, but actually find against their hypothesis.

While the Hurley and McDonough (1995) experiment did not support their hypothesis, it did emphasize the need of better analysis in the literature. While their experiment may very well be an accurate model of true betting behavior, it only had 18 participants. The empirical analyses also tend to focus on smaller data sets, limiting themselves to horse racing at a small selection of tracks.

This paper will allow for deeper analysis of the favorite-longshot bias as we improve on the existing literature in several key ways. First, we utilize a dataset containing data on more than 1.3 million betting events across a variety of sports, compared to the existing literature's focus on horse racing markets and thus provides a broader look at sports markets and a more robust data set with many more observations. Second, this paper is different in that there is no bookmaker involved. As Betfair is an exchange, existing arguments may need to be updated. For instance, Shin's work relies on bookmaker's setting profitable prices. In an exchange, where bookmakers don't play a role in setting prices, this argument will be less likely to explain the bias. Further, transaction costs on an exchange are significantly lower than traditional sportsbooks, so the use of an exchange can further test how easy access to information impacts the favorite-longshot bias. Finally, the use of a betting exchange simply provides a more accurate reflection of market activity than previous works do. An online exchange can be accessed by people from all around the world, and is not limited to an analysis of the people who physically show up to a racing track. Ultimately, existing literature has been unable to concretely explain why the favorite-longshot bias exits, and this paper has an opportunity to add to the analysis from both a behavioral and institutional view through an in depth empirical review.

3 Data

The data that we will be utilizing in this paper is a cross-section of bets placed on the betting exchange 'Betfair' over the course of one week in April, 2014. Each row of data represents a wager for a certain event at a given price (odds). It also includes how many individual people made each wager, as well as the total volume traded. Thus, it does not count each individual's bet separately, but rather aggregates all bets placed on one event at one price into one row. The data has a sample size of about 1.3 million and for the purposes of the analysis, we will operate under an assumption that this one week of data is a representative sample of Betfair's exchange. The primary variables we will be utilizing are ODDS, WIN_FLAG, NUMBER_BETS, VOLUME_MATCHED, IN_PLAY, and SPORTS_ID. ODDS is represented in European style odds format, such that the value represents the amount a \$1 wager would win plus the original investment (i.e. ODDS of 1.5 means that a \$1 wager would win \$0.50 as well as return the original \$1 invested). WIN_FLAG is a binary value that is 1 when the wager won, and 0 otherwise.

NUMBER_BETS is the total number of unique users that made the bet, and VOLUME_MATCHED is the total volume traded (bought or sold). IN_PLAY has a value of 0 when the bet was taken before the start of an event, and 1 when taken during the event as a 'live-bet'. Finally, SPORTS_ID is a unique identifier for each sport. For the purposes of analysis of efficiency, ODDS will be converted to a percentage form

$$\text{PERCENT\_CHANCE} = 1/\text{ODDS}\ , \tag{2}$$

and this will serve as an independent variable while WIN_FLAG will serve as a dependent variable in our initial test of efficiency.

To examine deviations from expectation, we calculate the returns on a \$1 investment:

$$r = \frac{\text{WIN\_FLAG} - \text{PERCENT\_CHANCE}}{\text{PERCENT\_CHANCE}}\ . \tag{3}$$

In a perfectly efficient market, the returns should be zero on average,

$$\frac{1}{n}\sum_{i=1}^{n} r_i = 0 \ . \tag{4}$$

These returns will be the dependent variables with NUMBER_BETS, VOLUME_MATCHED, and IN_PLAY as independent variables. SPORTS_ID will be used as a control to consider variations across markets for individual sports. Descriptive statistics of the data are presented in Table 7 of the Appendix.

Analysis is conducted in R, and tables are displayed with the assistance of Hlavak (2018)

4 Methodology

This paper will first test the original hypothesis of efficiency, followed by an analysis of deviations from expectation that occur. The model we use to test for efficiency follows Sauer (1998) and is of the form:

$$PW = \alpha H + \beta PC + \epsilon \tag{5}$$

where H is a vector of ones, PW refers to observed proportion of wins, PC refers to the percentage chance given by the odds in the market and ϵ is an error term. The joint null hypothesis is that $\alpha = 0$ and $\beta = 1$.

In practice, because the data refers to one realization of an event, we cannot gather the true proportion of wins from one data point. Similar to the procedure in Tompkins et al. (2003), we choose to create pools of 75 bets, all having the same ODDS, SPORTS_ID, and IN_PLAY values. We then calculate the expected proportion of wins PC for each pool, and can compare to the realized proportion of wins in the data set PW. Descriptive statistics of the pooled bets are presented in Table 8 of the Appendix.

A slope of the regression line that is different from one ($\beta \neq 1$) would indicate a bias of some sort. $\beta > 1$ would indicate that the favorites are undervalued while the longshots are overvalued, while $\beta < 1$ would indicate the reverse.

The plot in Figure 1 shows a clear linear trend in the pooled data, and as a result suggests that the Sauer model we use to test efficiency appears to be a reasonable one.

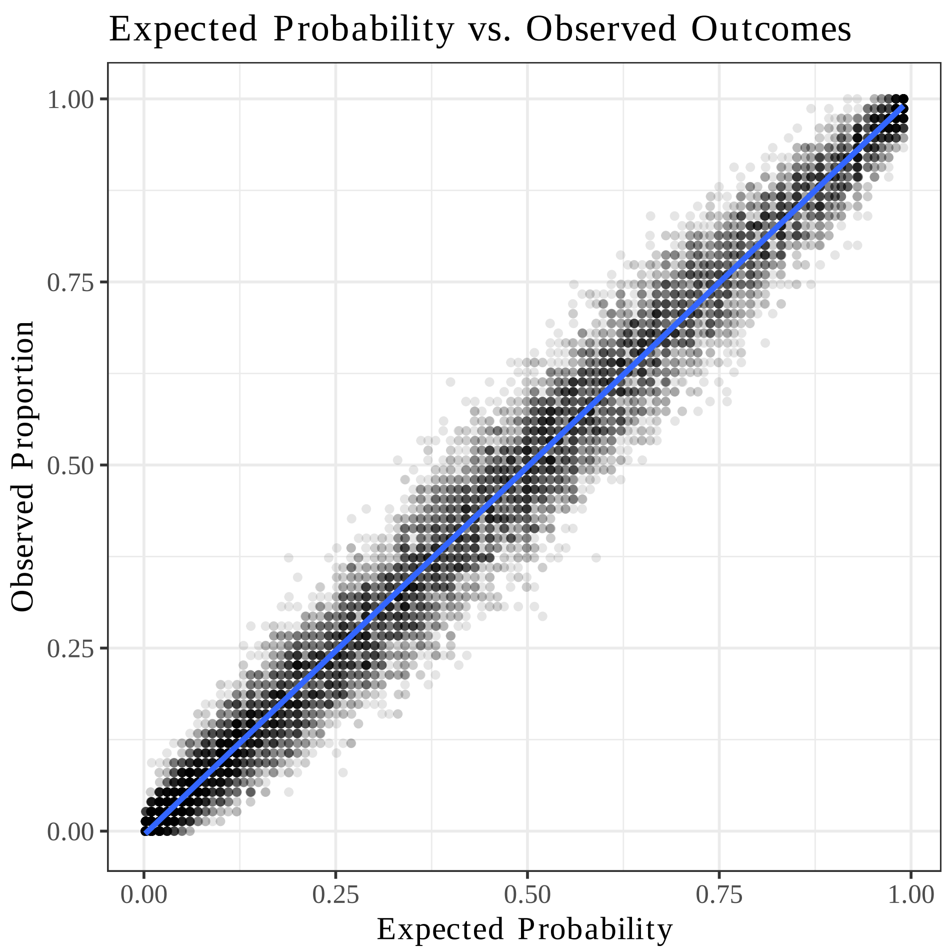

Figure 1: Expected proportion versus true proportion plotted for each group of pooled bets.

5 Initial Results

We test the original model ($PW = \alpha H + \beta PC + \epsilon$) and get the results displayed in Table 1. From the initial regression, despite the fact that the market is close to efficiency (Figure 1 shows a linear relationship), we reject the original null hypothesis that $\alpha = 0$ and $\beta = 1$.

Following a realization that the data looks linear, and is nearly efficient, it becomes the mission of this paper to determine where the inefficiencies lie, and why deviations (or bias) from the expected outcomes may exist. As $\beta > 1$, the data does appear to support the existence of the favorite-longshot bias. This bias in the data is also visualized in Figure 2. The plot shows the mean returns ($\bar{r}$) for each percentage point implied by the odds. These average returns (deviations) seem to have a linear and positive trend, despite some noise. As such, the original null hypothesis is rejected, and we conclude that the favorite-longshot bias is in fact present in the betting exchanges, seemingly in line with the vast majority of the literature on the subject. This bias now becomes the primary focus of the analysis in the remainder of this paper.

Table 1

	Dependent variable:
	PW
PC	1.006^{***}
	(0.001)
Constant	-0.005^{***}
	(0.001)
Observations	13,231
R^2	0.976
Adjusted R^2	0.976
Note:	*p<0.1; **p<0.05; ***p<0.01

6 Information Model

6.1 Approach

In order to examine the information approach of the favorite-longshot bias we extend our measure of deviations from the expected value of wins (returns) to the pooled data. The expected wins of a pool of data is calculated by taking the mean of the percent chances of

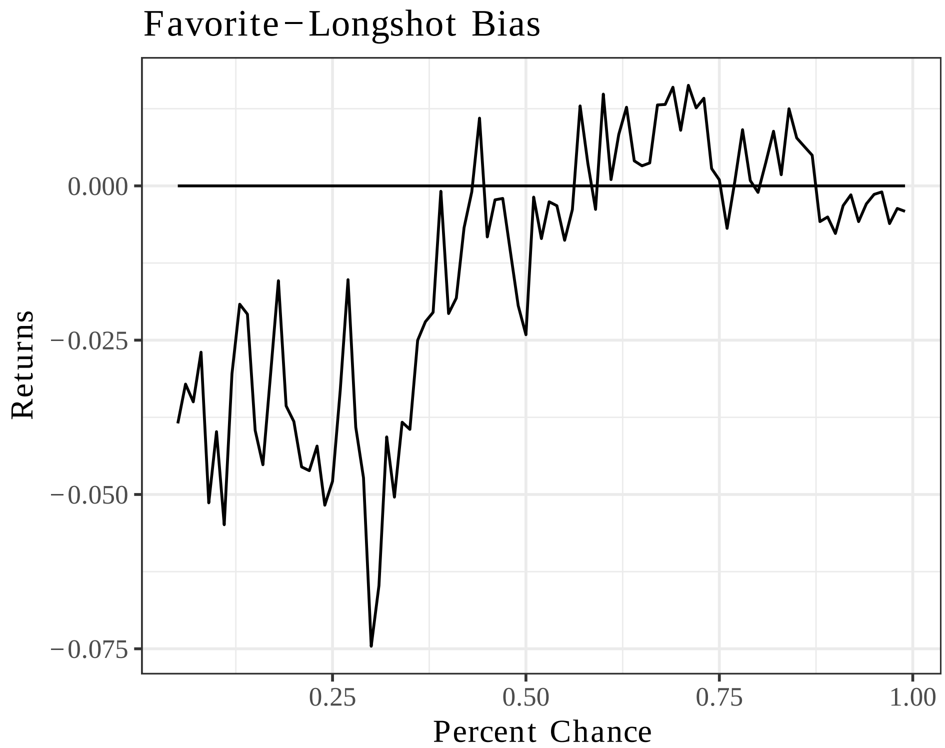

Figure 2: Returns plotted over market probability (PC).

each bet in the pool, as each bet has equal weight. The true fraction of wins comes from the number of observed wins divided by number of bets in the pool, such that our new equation for returns is as follows:

$$r = \frac{frac\_wins - expected\_wins}{expected\_wins} \tag{6}$$

We can now test the factors that cause changes in the value of returns. Hurley and McDonough (1995) hypothesized that the favorite-longshot bias's existence is due to an incomplete set of information for bettors. They argue that as information becomes accesible, this bias decreases. As they did not have data with which they could observe and test the effect of information, they conducted their own experiment with a small sample of bettors. However, an experiment such as this cannot be as good of a marker of the workings of a market as the market itself. The study found against their hypothesis, and that ultimately information is not a contributor to this bias.

Following Hurley and McDonough (1995) paper, the information theory has become a popular metric in subsequent work. Sobel and Raines (2003) modeled information at a sports betting track by comparing the number of bettors, suggesting that more bettors means more casual bettors, and thus a less informed betting pool. Their results found that this information effect is real. Meanwhile Smith et al. (2006) suggests that bets with more trading volume are more informed and also found similar results.

As such, we test the effects of volume and number of bets on returns. We also include our own metric of information, that of bets being placed after the event begins versus those placed before. We have *in_play* serving as an indicator variable representing whether or not a given pool of bets was placed during the course of a sporting event. We suggest that on average a bet taken after a sporting event begins is more informed than a bet taken before the sporting event begins. This is due to the fact thast as the game begins, any injuries, special abilities, etc. of a participant become apparent, allowing for more information available to bettors simply as a function of time. Thus we expect a pool of in-game bets to be more informed than a pool of pre-event bets.

Before conducting our tests on the causes of deviations, we justify the use of the *in_play* indicator by conducting the following regression on the pooled data:

$$PW = \alpha H + \beta_E PC + \beta_{IP} in\_play + \epsilon \tag{7}$$

The results of this regression are available in Table 10 of the Appendix, both in its original form and with sport fixed effects. In either case, we see a significant coefficient on *in_play*, suggesting that the fact that a bet was placed during the game is informative to predicting the outcome of the game, and thus we determine that *in_play* is a valid metric of level of information.

We conduct the following tests:

$$r = \alpha H + \beta_E PC + \epsilon \tag{8}$$

$$r = \alpha H + \beta_E PC + \beta_V Volume + \beta_{bias\_vol}(PC * Volume) + \epsilon \tag{9}$$

$$r = \alpha H + \beta_E PC + \beta_N num\_bets + \beta_{bias\_num}(PC * num\_bets) + \epsilon \tag{10}$$

$$r = \alpha H + \beta_E PC + \beta_{IP} in\_play + \beta_{bias\_ip}(PC * in\_play) + \epsilon \tag{11}$$

where H is a vector of ones, $r = \frac{PW-PC}{PC}$ and ϵ is an error term. We also add a fixed effects model, $r = \alpha_i + \beta_t X_t + \epsilon$ for $i = (1, ..., n)$ where i represents each SPORTS_ID, and for $t = (1, ..., m)$ where m is the number of independent variables in the regression for the analysis of effects on the bias. This is used to control for any effects that may be attributed to one sport but not another. This follows Cain et al. (2003) which suggests some sports have differing degrees of the favorite-longshot bias.

In order to see the effects of these factors on the bias we choose to examine how these factors affect the slope of this favorite-longshot relationship. As we have shown the existence of a favorite-longshot bias above, we expect β_E, the coefficient on expected probability to be positive in equation (8). Meanwhile, to test the impact on the favorite-longshot bias, we test how the relationship between market probability and returns changes based on these factors, hence the interaction terms in equations (9) - (11).

6.2 Results

The results of the regressions from the previous section are visible in Tables 2 and 3. Regressions were conducted only on events with a market probability greater than 10% as lower probabilities had high noise in returns. The regression in Table 2 verifies the existence of the favorite-longshot bias, as there is a positive coefficient on PC, indicating a trend as shown in Figure 2.

Regressions (1)-(3) in Table 3 allow the analysis of the severity of the favorite-longshot bias. However, the interaction terms for volume and market probability, as well as *num_bets* and market probability show absolutely no significance. This suggests that perhaps the conclusions of Sobel and Raines (2003) and Smith et al. (2006) simply do not scale, and no longer apply when considering a large betting exchange, thus failing to explain the whole of the bias. Meanwhile, The interaction term for *in_play* and market probability actually suggests that more informed bets have a stronger case of the favorite-longshot bias as being an in play bet makes average returns increase by 0.080 more per percentage point increase in expected probability than a pre-event bet. The addition of sport fixed effects does not seem to change the significance of any of these results.

As a check on the robustness of the analysis, we also consider the non-pooled, original data, specifically events that have odds both pre-game and in-game. The descriptive statistics for this data are available in Table 9 of the Appendix. Here we conduct the following analysis:

$$|r| = \alpha H + \beta_E PC + \beta_{IP} in\_play + \epsilon \tag{12}$$

$$r = \alpha H + \beta_E PC + \beta_{IP} in\_play + \beta_{bias\_ip}(PC * in\_play) + \epsilon \tag{13}$$

This examination should show on an individual event basis, whether or not the in-game odds will yield less deviation from the expected value and whether or not it will lower the degree of bias. We also use fixed effects for each sporting event, to account for any added features that may be attributed to a particular event. As shown in Table 4, and much like our analysis of equation (7), it does appear that the *in_play* factor lowers overall deviation. However, the interaction term acts to increase the level of deviation, and thus we still are not able to conclude that more informed bets have a lower impact of the favorite-longshot bias than do less informed bets.

Ultimately, these results resoundly reject the information cost explanation for the favorite-longshot bias. Using three separate metrics for information this conclusion is achieved, so while the information costs may be an accurate representation of horse racing tracks and smaller markets, it is unlikely to have explanatory power outside of these niche markets.

Table 2

	Dependent variable:
	returns
PC	0.058***
	(0.006)
Constant	−0.043***
	(0.003)
Observations	10,684
R^2	0.009
Adjusted R^2	0.009
Note:	*p<0.1; **p<0.05; ***p<0.01

Table 3

	Dependent variable:					
	returns					
	OLS			*felm*		
	(1)	(2)	(3)	(4)	(5)	(6)
PC	0.058***	0.061***	−0.004	0.056***	0.046***	−0.002
	(0.006)	(0.012)	(0.012)	(0.006)	(0.012)	(0.012)
volume	0.00000			0.00001***		
	(0.00000)			(0.00000)		
num_bets		0.001			0.002**	
		(0.001)			(0.001)	
ip			−0.064***			−0.063***
			(0.008)			(0.008)
interact	−0.00000	−0.001	0.080***	−0.00001***	0.0005	0.081***
	(0.00000)	(0.001)	(0.014)	(0.00000)	(0.001)	(0.014)
Constant	−0.044***	−0.049***	0.006			
	(0.003)	(0.007)	(0.007)			
Observations	10,684	10,684	10,684	10,684	10,684	10,684
R^2	0.009	0.009	0.015	0.018	0.018	0.021
Adjusted R^2	0.009	0.009	0.015	0.017	0.017	0.019
Note:						*p<0.1; **p<0.05; ***p<0.01

Table 4

	Dependent variable:		
	abs_returns	returns	
	OLS	*OLS*	*felm*
PC	-1.708^{***}	0.215^{***}	0.204^{***}
	(0.009)	(0.021)	(0.024)
IP	-0.094^{***}	-0.384^{***}	-0.394^{***}
	(0.006)	(0.022)	(0.022)
interact		0.477^{***}	0.493^{***}
		(0.031)	(0.031)
Constant	1.804^{***}	-0.262^{***}	
	(0.007)	(0.015)	
Observations	44,842	44,842	44,842
R^2	0.429	0.024	0.065
Adjusted R^2	0.429	0.024	0.039
Note:			*p<0.1; **p<0.05; ***p<0.01

6.3 Discussion

The results appear to be counter to the literature's expectations in regards to the effect of a more informed wager. It appears that the analysis from Sobel and Raines (2003) and Smith et al. (2006) does not hold in the context of a betting exchange, and that our own metric of more informed bets actually causes an increase in the bias. This leads to a conclusion that information costs may not be the cause of a bias. Transaction costs, too, seem unimpactful as all bets have the same commission charges, yet bets that were more informed had a greater bias.

In fact, applying the original model ($PW = \alpha H + \beta PC + \epsilon$) on only bets that occured pre-event, we find that we cannot reject the null hypothesis that the predicted probabilities are unbiased estimators of the true proportion of wins ($\alpha = 0$ and $\beta = 1$) at the 95 percent confidence level. These results are in Table 5 with 95% confidence intervals displayed.

Consequentially, we find that the pre-event bets on Betfair's service are in fact mostly efficient, and any inefficiencies that occur, including the favorite-longshot bias, occur as a

Table 5

	Dependent variable:
	PW
PC	0.997***
	(0.992, 1.002)
Constant	0.002*
	(−0.0004, 0.005)
Observations	2,964
R^2	0.980
Adjusted R^2	0.980
Note:	*p<0.1; **p<0.05; ***p<0.01

result of some feature of the in the in-game betting.

Perhaps, there are different forces that are at play that cause the bias instead. With the more informed bets having a greater bias, it begs the question of how exactly information is being utilized. Piccoli et al. (2017) suggests that markets tend to overreact when facing news and events. In play bets would be more susceptible to this information overreaction as there is simply more movement and change that could affect the outcome of a game once it begins, than before it starts. It is possible for instance, that when a team's key player becomes injured the team's odds of winning would decrease by more than they should, creating some deviation from the expectation. It is not immediately clear why these deviations would be in the form of a favorite-longshot bias, but it does raise interesting questions on how bettors absorb information as it is changing.

Beyond this, it is clear now that many conclusions on the causes of the favorite-longshot bias seem doubtful in the context of a betting exchange. Conclusions relating to sportsbook behavior don't seem to hold when there is no one acting as a sportsbook, yet we still find evidence of a bias without one. This study expands on the previous use of small samples of bets from horse racing markets and utilizes many sports with many more data points, and finds that many of the same trends still exist, therefore explanations that relied on features specific to those markets simply don't hold.

Ultimately, this paper finds that markets do tend to be quite efficient, and even have no significant deviations from expectation for pre-event bets. These conclusions can help us understand broader market structure for other financial markets as well. For instance, Tompkins et al. (2003), find evidence of this same favorite-longshot bias in some options markets. The findings from this paper, suggesting that it is not transaction/information costs, or risk preferences that affect prices that cause the bias, can then be extrapolated to those markets as well, and suggest that deviations from expectations in options markets are not due to a lack of information.

7 Further Analysis

We see from the results in this paper that markets behave differently for in-play and pre-event wagers, and that deviations from expectations are not remedied with greater information availability. In order to understand why such a difference exists, and why the in-game bets exhibit a bias, we consider the population of bettors. Despite the fact that more information is available, the in-game bets are exhibiting a bias that the less informed pre-game bets does not. This leads to speculation that perhaps those that trade in the pre-event wagers are more informed or professionals, while in-game bettors are not. This follows the analysis from Osborne (1962) which suggests that depending on day of the week there was a remarkable difference in the number of odd lots vs. round lots of stock traded. Round lots are more likely to be traded by professionals, while orders of odd lots are likely to be made by non-professional traders. Perhaps in the world of sports betting, the money of professionals is on pre-event bets, with in-game betting being left to the non-professional bettor. In order to test this hypothesis we consider bet size. By examining the average bet size of wagers on each individual event, we compare the populations making pre-event and in-play bets.

After conducting a Wilcoxon Rank Sum Test, we reject the hypothesis that thes bets come from the same distribution at a 99% confidence level, in support of the belief that these populations of bettors have some inherent differences. Specifically we find that the bet sizes are larger pre-event than in-game, in line with conclusions from Osborne (1962) on professional traders having different behavior than non-professionals when it comes to the size of their trades. This could also suggest that although in-game bettors may have access to more information, they may not necessarily be using it properly, leading them to either under or overreact to new information, as was found in De Bondt and Thaler (1985) in their analysis of the impacts of dramatic news events on stock prices and the Overreaction Hypothesis.

By considering only events that had trading before and during the event, we examine the difference in implied probability from the initial pre-event odds to the in-play odds. A high difference in probability is likely to occur due to some drastic event such as the injury of a key player, while mundane updates in score would result in negligible movement of probabilities. we thus use these changes in odds as proxies for the value that traders place on new information, hypothesizing that larger changes in percieved probabilities will result in returns that deviate more from expectation, thus serving as a contributor to the favorite-longshot bias.

For the analysis we consider all large changes in probability (shifts greater than 15 percentage points) as smaller changes can likely be attributed to noise, and conduct a regression similar to the ones done to test the information model as presented in equation (**??**), where H is a vector of ones, $r = \frac{PW-PC}{PC}$ and ϵ is an error term:

$$r = \alpha H + \beta_E PC_{IP} + \beta_\delta \delta PC + \beta_{bias\_\delta PC}(PC_{IP} * \delta PC) + \epsilon \tag{14}$$

This regression is conducted twice, once for all wagers with positive changes in betting odds, and once for all wagers with negative changes in betting odds. The results are shown in Table 6, with (1) examining positive odds shifts, and (2) examining negative odds shifts.

These results show that for large positive odds increases, there is a significant positive increase in the relationship between the theeoretical probability and returns, a sign of a

Table 6

	Dependent variable:	
	ip_returns	
	(1)	(2)
IP_PC	0.508***	1.856***
	(0.104)	(0.485)
odds_change	−1.364***	−1.642***
	(0.296)	(0.472)
interact	1.301***	4.335**
	(0.335)	(1.714)
Constant	−0.419***	−0.938***
	(0.084)	(0.160)
Observations	4,370	2,444
R^2	0.064	0.009
Adjusted R^2	0.063	0.008
Note:	*p<0.1; **p<0.05; ***p<0.01	

strengthening favorite bias. Meanwhile, large negative changes, result in a sharp decline in the returns, showing a strong longshot bias. Overall, we see evidence in line with the findings of De Bondt and Thaler (1985), as it appears that higher changes in probability tend to have a larger amount of bias in both the positive and negative case. Further, as in Piccoli et al. (2017), the degree to which this occurs seems to be higher for negative news (decrease in probability) than it does for positive news (increase in probability), as the coefficient on the interaction between the implied probability and shift in odds is much higher for the negative shifts (2).

7.1 Prospect Theory

Thus far, we have shown that although pre-event bets tend to be rather efficient, and do not exhibit a bias, the more informed in-play bets tend to overvalue low probability events and undervalue high probability events. We have also seen that these in play bets tend to be made in smaller size, suggesting that they are not made by professionals, and that large news events tend to make the degree of bias higher. Next, we attempt to understand the reason why such a phenomenon might occur for the lower volume traders.

In order to examine how these individuals make decisions, we consider prospect theory which examines risky prospects in an experimental setting. In contrast to expected utility theory which suggests that the utility of a prospect is equivalent to the sum of the utilities of its potential outcomes multiplied by their respective probabilities of occuring, Tversky and Kahneman (1992) suggests that the utility of a risky prospect should be a function of the gain or loss from that prospect and a respective decision weight. They also provide updates on the original prospect theory literature by suggesting a cumulative prospect theory in which $V(f)$, or the value of a prospect f is given by

$$V(f) = \sum_{-m}^{n} \pi_i v(x_i) \ . \tag{15}$$

Further, they propose that for positive prospects, the value function is of the form

$$v(x) = x^{\alpha} \tag{16}$$

where $0 \leq \alpha \leq 1$, and x is the outcome of a prospect. The weighting function is of the form

$$w(p) = \frac{p^{\gamma}}{(p^{\gamma} + (1-p)^{\gamma})^{\frac{1}{\gamma}}} \ . \tag{17}$$

Finally, they conduct an experiment in which members of the study were asked a series of questions, choosing between prospects and alternative guarantees of gain or loss. Tversky and Kahneman (1992) then estimate the weighting function as c/x, where c is the certainty equivalent of a prospect, and x is its non-zero outcome. As a result they conclude that the weighting function follows an inverted S-Shape, where individuals tend to overweight low probabilities, and underweight high probabilities. Expanding on this work, Tversky and Fox (1995), shows that this same analysis applies not only to risky prospects in which probability of an outcome is known, but also uncertain prospects (such as sports betting or investing in stocks), when using a judged probability.

Following this methodology, we consider the in-play bets. As these bettors place their wagers during the match, the pre-event odds are available and serve as judged probabilities. By our earlier findings that the pre-event bets are rather efficient, these judged probabilities are likely good estimates of the true probabilities. As we are using a betting exchange, the odds at which a trade takes place represents the highest value for which an individual would exchange a guaranteed amount for a prospect, as well as the lowest amount for which another individual would trade a prospect for a guaranteed return. As such, our odds themselves represent the certainty equivalent. Importantly, since we assume a power value function in equation (16), we suggest that the certainty equivalent is a linear function of the prize of the prospect, just as in Tversky and Kahneman (1992). That is, for a certainty equivalent function $C(x)$

$$C(\lambda x) = \lambda C(X) = \lambda c \tag{18}$$

for some constant λ. Thus, the size of a bet has no impact on c aside from scaling it, so we are able to treat all prospects in our data the same regardless of bet size (c/x is not dependent on bet size). In accordance with Tversky and Kahneman (1992), we model our weighting function as c/x. As we have converted all odds to percentage form, the outcome of each prospect is either 0 or 1, so $x = 1$ and thus our weighting function is represented by c, where c is equal to the in-play odds.

We plot our weighting function against the judged probabilities in Figure 3, and use a non-linear regression to fit the model in equation (17). This results in a fitted value $\gamma = 0.6991$, very similar to results from Tversky and Kahneman (1992). Additionally, this plot looks almost identical to their analogs presented in Tversky and Kahneman (1992) and Tversky and Fox (1995). This is significant as it expands upon the experimental studies done on small samples of students in both of those papers, by providing observational data from over 1.3 million betting events. Thus, we arrive at a similar conclusion, that individuals

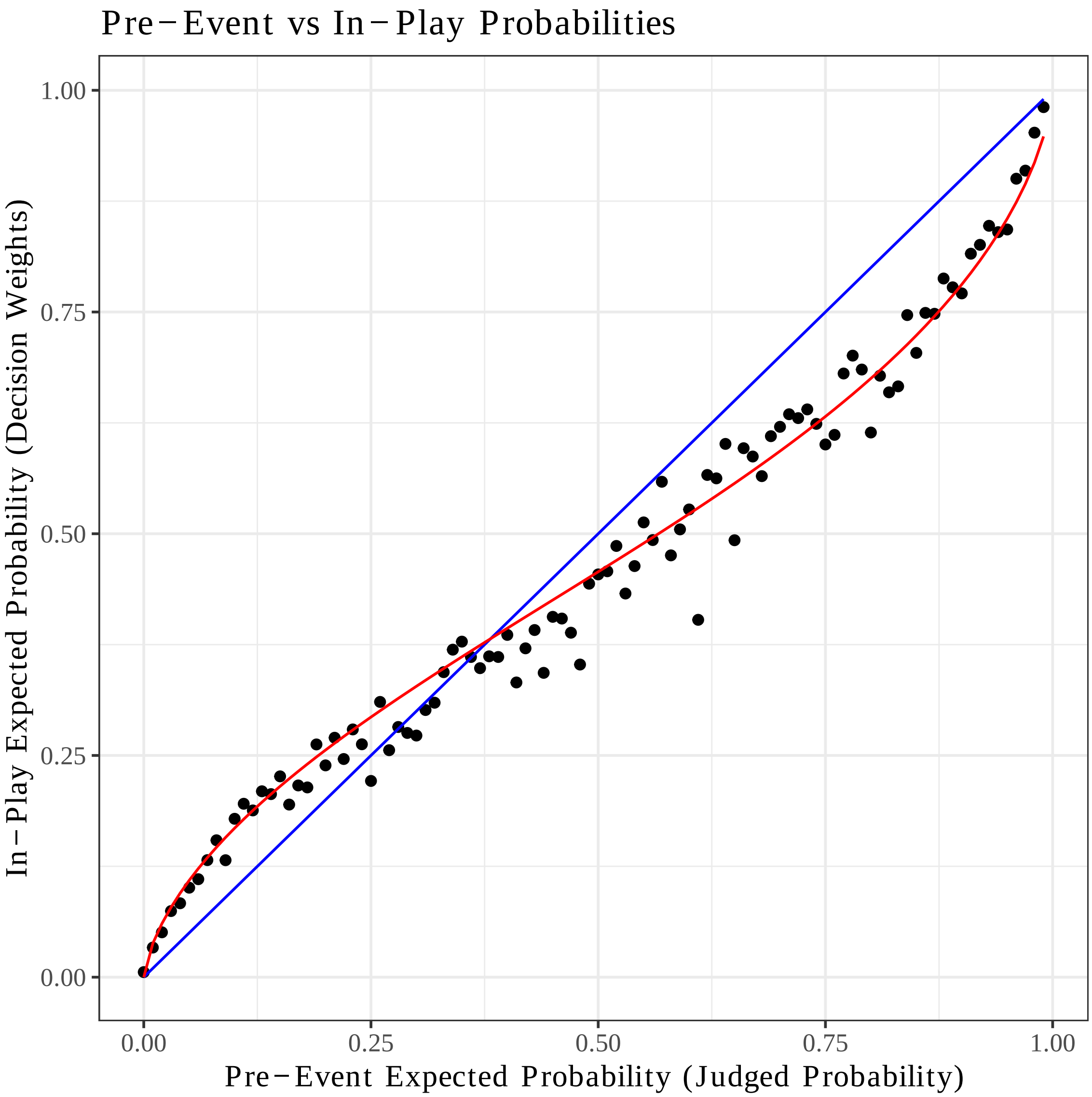

Figure 3: Mean in-play implied probabilities plotted against pre-event implied probabilities.

tend to have a weighting function that is not linear with probability, but rather one that has an inverted S-Shape, but using a vast data set that expands the previous work.

This also provides potential explanation for our findings on the favorite-longshot bias. We discover that the pre-event bets do in fact have odds-implied probabilities that are linear with true probabilities and that their differences are indistinguishable from zero. Based on the above discussion of prospect theory, this is equivalent to having a linear weighting function. We also find, that betting size is statistically larger in pre-event bets than in in-play bets, suggesting underlying differences in betting participants. As a result, we find that the weighting function for pre-event bettors, which by their large betting size we suggest are professionals, is linear, with decision weights as explained by expected utility theory. Meanwhile for non-professionals, which we suggest populate the in-play betting field, the weighting function is one that underweights high probability events, and overweights low probability events, causing the favorite-longshot bias that we have observed.

8 Concluding Remarks

Sports betting markets provide for a clean way to view the manner in which markets price events with a finite amount of outcomes. These outcomes can be measured and compared to the prices at which they were traded. It is this ease of analysis that makes sports markets so interesting for many looking to observe the efficiency of markets. This work follows the likes of many others, primarily those in horse racing sportsbooks, and expands by utilizing a betting exchange with a variety of sports and bets available to anyone in the world.

This paper seeks to discover if these markets are efficient. As a whole it finds that markets do tend to be close to the true outcomes in their predictions, but do have significant deviations from expectations using the model from Sauer (1998). Upon further review of these deviations, we find evidence of the so called favorite-longshot bias, in which favorites are under priced relative to their true outcomes, and longshots are overpriced. That is to say that consistently betting on favorites will yield positive returns, and betting on longshots will yield negative returns, a clear violation of efficiency.

We look to find the root of the deviation by studying how individuals gain and utilize information. On information costs and access, we test the impact of volume and number of bettors, similar to Sobel and Raines (2003) and find that the factors are insignificant to the bias that is observed. We also use our own proxy for information, that of a flag for bets taken after games begin, and still find that the information has no significant impact in decreasing the bias. Instead, our results show that in play bets are more biased than pre-event bets, with pre-event bets being rather efficient. In search of an explanation why, this paper considers differences between the bettors that participate in the types of bets, and finds that overwhelmingly, the bet size in pre-event bets is larger than in in-game bets, suggesting that the pre game bettors are more likely to be professionals than the in-game bettors. Further, we see that as betting odds change during an event, larger changes lead to a strengthening of the bias, suggesting that adjustments to new and influential information, have large contributions to the favorite-longshot bias.

Finally, we consider prospect theory and find results consistent with Tversky and Fox (1995), suggesting that behavioral reasons in the form of a non-linear weighting function used in the calculation of individual's utility may be the guiding principle for the cause of the favorite-longshot bias. Ultimately, we reject the notion that the favorite-longshot bias is caused by a lack of information available to bettors, and instead conclude that individuals do not necessarily weight their decisions rationally based on the information they absorb, and that this phenomenon is a likely reason for the bias we observe.

This paper was written during the COVID-19 pandemic, a time of great uncertainty in markets, and in people's lives. The manner in which individuals have difficulty handling uncertainty has been on full display, whether it be politicians having difficulties closing and reopening economies, consumers hoarding toilet paper, or markets behaving in seemingly erratic ways. This pandemic, while certainly tragic, has been quite an opportunity to see irrationality at work. Ultimately, whether it be in sports betting markets or otherwise, decisions made by individuals seem to not be fully reflective of the information that guides them. In this paper, we find that markets are in fact relatively efficient, yet they do exhibit a significant favorite-longshot bias, which can be attributed at least in part to the mishandling of significant probability altering information.

References

Cain, M., Law, D., and Peel, D. (2003). The favourite-longshot bias, bookmaker margins and insider trading in a variety of betting markets. *Bulletin of Economic Research*, 55(3):263–273.

De Bondt, W. F. M. and Thaler, R. H. (1985). Does the Stock Market Overreact? *The Journal of Finance*, 40(3):793–805.

Hlavak, M. (2018). stargazer: Well-Formatted Regression and Summary Statistics Tables. R package version 5.2.1.

Hurley, W. and McDonough, L. (1995). American Economic Association A Note on the Hayek Hypothesis and the Favorite-Longshot Bias in Parimutuel Betting. 85(4):949–955.

Osborne, M. F. M. (1962). Periodic Structure in the Brownian Motion of Stock Prices. *Operations Research*, 10(3):345–379.

Piccoli, P., Chaudhury, M., Souza, A., and da Silva, W. V. (2017). Stock overreaction to extreme market events. *North American Journal of Economics and Finance*, 41(514):97–111.

Quandt, R. E. . (1986). Betting and Equilibrium. *The Quarterly Journal of Economics*, 101(1):201–208.

Sauer, R. D. (1998). The Economics of Wagering Markets. *Journal of Economic Literature*, 36(4):2021–2064.

Shin, H. S. (1992). Prices of State Contingent Claims with Insider Traders, and the Favourite-Longshot Bias. *The Economic Journal*, 102(411):426.

Smith, M. A., Paton, D., and Williams, L. V. (2006). Market efficiency in person-to-person betting. *Economica*, 73(292):673–689.

Sobel, R. S. and Raines, S. T. (2003). An examination of the empirical derivatives of the favourite-longshot bias in racetrack betting. *Applied Economics*, 35(4):371–385.

Thaler, R. H. and Ziemba, W. T. (1988). Anomalies: Parimutuel Betting Markets: Racetracks and Lotteries. *Journal of Economic Perspectives*, 2(2):161–174.

Tompkins, R. G., Ziemba, W. T., and Hodges, S. D. (2003). The Favorite-Longshot Bias in s&p 500 and FTSE 100 Index Futures Options: The Return to Bets and the Cost Of Insurance. *Handbook of Sports and Lottery Markets*, 10:161–180.

Tversky, A. and Fox, C. R. (1995). Weighing risk and uncertainty. *Psychological Review*, 102(2):269–283.

Tversky, A. and Kahneman, D. (1992). Advances in prospect theory: Cumulative representation of uncertainty. *Journal of Risk and Uncertainty*, 5(4):297–323.

Appendix

Table 7

Statistic	N	Mean	Median	Min	Max	St. Dev.
ODDS	1,306,746	21.322	2.980	1.010	1,000.000	91.004
WIN_FLAG	1,306,747	0.379	0.000	0.000	2.050	0.485
NUMBER_BETS	1,306,750	277.244	160	1	784	190.923
VOLUME_MATCHED	1,306,750	48,336.890	51,170	1	107,801	31,941.740
IN_PLAY	1,306,750	0.618	1	0	1	0.486
SPORTS_ID	1,306,750	16,685.670	1	1	26,420,387	541,842.100
PERCENT_CHANCE	1,306,746	0.382	0.336	0.001	0.990	0.290

Table 8

Statistic	N	Mean	Median	Min	Max	St. Dev.
deviations	13,231	−0.003	−0.002	−0.226	0.214	0.047
returns	13,231	0.007	−0.01	−1	12	0.584
PC	13,231	0.419	0.391	0.002	0.990	0.296
PW	13,231	0.416	0.4	0	1	0.301
volume	13,231	698.667	151.628	0.511	71,024.380	2,629.748
num_bets	13,231	7.565	6.067	1.013	109.613	5.442
ip	13,231	0.776	1	0	1	0.417
sport	13,231	323.634	1	1	998,917	12,329.180
abs_returns	13,231	0.231	0.1	0	12	0.536

Table 9

Statistic	N	Mean	Median	Min	Max	St. Dev.
X	68,476	49,113.380	49,648.5	10	86,507	21,238.630
ODDS	68,476	68.652	3.8	1	1,000	207.210
IP	68,476	0.500	0.5	0	1	0.500
PW	68,476	0.386	0	0	1	0.487
PC	68,476	0.424	0.267	0.001	0.990	0.393
deviations	68,476	−0.038	−0.011	−0.990	0.999	0.285
returns	68,476	−0.227	−1	−1	989	7.080
abs_returns	68,476	1.001	1	0	989	7.013

Table 10

	Dependent variable:	
	PW	
	OLS	*felm*
PC	1.013***	1.014***
	(1.009, 1.017)	(1.010, 1.018)
ip	−0.006***	−0.006***
	(−0.009, −0.004)	(−0.009, −0.004)
Constant	−0.005***	
	(−0.008, −0.002)	
Observations	10,684	10,684
R^2	0.964	0.964
Adjusted R^2	0.964	0.964
Note:		*p<0.1; **p<0.05; ***p<0.01