Have Betting Exchanges Corrupted Horse Racing?*

Alasdair Brown[†]

Fuyu Yang[‡]

University of East Anglia

University of East Anglia

May 19, 2015

†Corresponding Author. School of Economics, University of East Anglia, Norwich NR4 7TJ, U.K.. Email: alasdair.brown@uea.ac.uk Tel: 44 1603 591131

‡School of Economics, University of East Anglia, Norwich NR4 7TJ, U.K.. Email: fuyu.yang@uea.ac.uk Tel: 44 1603 591058

Abstract

Betting exchanges allow punters to bet on a horse to lose a race. This, many argue, has opened up the sport to a new form of corruption, where races will be deliberately lost in order to profit from betting. We examine whether anecdotal evidence of the fixing of horses to lose - of which there are many examples - is indicative of wider corruption. Following a 'forensic economics' approach, we build an asymmetric information model of exchange betting, and take it to betting data on 9,560 races run in 2013/14. We find no evidence of the widespread corruption of horse racing by the betting exchanges.

1   Introduction

Betting exchanges, such as Betfair and Betdaq in the U.K., mimic the structure of financial exchanges, by allowing bettors to take long or short positions on bets. For example, bettors can 'back' a horse (bet on a horse to win a race) or 'lay' a horse (bet on a horse to lose). Punters who bet on horses to lose assume a role that, prior to the establishment of betting exchanges in the year 2000, had for centuries been the exclusive purview of bookmakers. By facilitating such 'person-to-person' betting, the exchanges have radically altered the nature of betting in the U.K.. Smith *et al.* (2006) find that transaction costs are much lower on the exchanges than they are with bookmakers. Betfair, the largest of the exchanges, now hosts more than 1.2 billion bets a year across a range of sports.[1]

This upheaval in the betting industry has not been without controversy. Bookmakers, regulators and others within the industry have argued that betting exchanges threaten the integrity of horse racing. The argument is that by allowing bettors to profit from a horse losing a race, exchanges encourage jockeys, trainers and owners to deliberately lose races. While fixing a horse to win a race is tricky - as it requires the cooperation of all other participants in the race - fixing a horse to lose only requires the actions of those connected to one horse. This

---

[1]http://corporate.betfair.com/about-us/a-betting-revolution.aspx

has caused many to doubt whether all horses are being run on their merits. As the trainer Mark Johnston said - in an article titled 'Does the existence of exchanges corrupt sport? - 'exchanges give an opportunity to cheat that wasn't there before their arrival'.[2] Concerns for the integrity of racing have also, at one time or another, been raised by regulators in Australia, France, Japan and California as an argument against the legalisation of betting exchanges (O'Connor (2014)).

There is plenty of anecdotal evidence to support the view that betting exchanges have facilitated the corruption of horse racing. In August 2004, Fergal Lynch deliberately stopped his horse, Bond City, in a race at Ripon. His actions were at the behest of the horse's owner Miles Rogers, who had laid the horse to lose on Betfair. Lynch was banned from racing for life, but recently returned to the track.[3] In 2010-11, the jockey Andrew Heffernan was part of a conspiracy to deliberately lose 9 all-weather races, and profit by betting on each horse to lose on the exchanges. This conspiracy, which also involved a number of professional footballers as co-conspirators, resulted in lengthy bans for all those involved.[4] In January 2012 Steven Gagan, a jump jockey, leapt off the favourite Kickahead while in the lead of a race at Musselburgh. Again, connections of this horse had laid the horse to lose on the exchanges.[5] Many more cases can be found on the British Horse Racing Authority's disciplinary hearings page.[6]

How widespread is this practice? Have betting exchanges, by facilitating lay betting and allowing for losers to profit, inadvertently corrupted horse racing? In this paper we analyse whether the examples of corruption described above are just isolated incidents, or signs of a wider problem in the sport.

Answering such a question is difficult as corruption, by its very nature, is hidden from view. Our approach is to develop an asymmetric information model of betting exchange

---

[2]'Does the existence of betting exchanges corrupt sport?', Mark Johnston and Mark Davies, *The Guardian*, 12th December 2007.

[3]'BHA treats punters as mugs by letting disgraced Fergal Lynch ride again', Greg Wood, *The Guardian*, 10th August 2014.

[4]'Chopra guilty of race-fixing scam', Chris McGrath, *The Telegraph*, 25th January 2013.

[5]'Riddle over the case of Steven Gagan, the jockey who deliberately fell off', Greg Wood, *The Guardian*, 24th August 2014.

[6]http://www.britishhorseracing.com/resource-centre/disciplinary-results/disciplinaryappeal-hearings/

activity and compare it with pricing data from Betfair. Our model is based on Glosten and Milgrom (1985), who had New York Stock Exchange dealer markets in mind, and Shin (1993), who analysed traditional U.K. bookmaker markets. We adapt these models to fit the characteristics of betting exchanges, where a binary asset is traded on a limit order book, and where punters can back a horse to win, or lay a horse to lose. Our set-up is simple. A Bayesian market-maker (who voluntarily provides quoted back and lay odds) trades bets with a population of informed bettors (who have fixed the outcome of the race) and uninformed noise bettors, who trade randomly. The market-maker quotes back and lay odds dependent on her unbiased prior estimation of the horse's win probability (which approximates the empirical probability), and dependent on her estimation of the relative proportion of informed and uninformed bettors in the population. If a back order is received, the market-maker revises up her beliefs of the horse's win probability (as long as there are some informed in the population), and if a lay order is received the market-maker revises down her estimate of the horse's win probability. This creates an adverse selection spread between back and lay quotes.

The empirical implications of the model are as follows. If there is adverse selection - i.e. if the market-maker can be 'adversely-selected' by bettors that have fixed a horse to win or lose - then back bet prices (odds) will overestimate the probability that a horse wins, and lay bet prices will underestimate the probability that a horse wins. This is because the market-maker recoups the losses to the informed with gains from the noise bettors. More importantly, if betting exchanges have corrupted horse racing by allowing for cheats to fix horses to lose, then lay bets should overestimate a horse's win probability by more than the back bets underestimate the horse's win probability. This is because there is greater adverse selection risk associated with lay bets than with back bets, as it is easier to fix a horse to lose than it is to fix a horse to win. If, on the other hand, back and lay bet prices deviate from the empirical win probabilities by equal amounts (i.e. adverse selection arises equally from backers and layers) - or both back and lay bets closely correspond with empirical win probabilities - then there is little to suggest that betting exchanges have corrupted horse racing.

We compare this model to pricing data from 9,560 U.K. horse races run in 2013 and 2014.

We use more than 225 million observations of the Betfair limit order book to construct data on the direction and volume of all bets placed in the 20 minute window before each race. This constitutes more than 80% of all betting volume and involves more than 60 million separate trades. Crucially, we divide up these trades into those initiated in the form of back orders, and those initiated in the form of lay orders. We then construct the trade-weighted average back and lay price for each horse in each race. This gives a very precise representation of the prices at which back and lay bets are executed. We find that both back and lay bet prices closely track the empirical win probabilities of horses. In other words, there does not appear to be an adverse selection component to prices in this market. This implies that bettors in this market believe that the events described earlier are isolated incidents, rather than a widespread phenomenon that has corrupted horse racing.

This paper falls into a type of research recently labelled 'forensic economics' (Zitzewitz (2012)). As with other examples of this work, we 'size the extent of an activity about which there had been only anecdotal evidence and provide insight into where it is more prevalent and why'.[7] While 'wrongdoers may cover their tracks well enough to fool traditional forensic investigators, ... they leave distortions in the data that economic analysis can detect.[8] An example of this approach is found in DellaVigna and Ferrera (2010). The authors examine the stock prices of arms companies in the aftermath of conflict intensifications. The aim is to identify whether a given company is involved in the illicit selling of arms. The idea is that while investigators may not be able to pin down, or definitively prove, illegal arms sales, financial market participants have good incentives to accurately predict whether, on the balance of probabilities, a given firm is involved. If the stock price spikes in the aftermath of a ramp-up in conflict, this suggests that the firm is, in some way, shepherding arms to the conflict region.

An example from sport is the work of Wolfers (2006) and Bernhardt and Heston (2010), who examine 'point-shaving' in NBA basketball. This is the practice by which heavily favoured teams deliberately win by just less than the quoted spread predicts. This allows the team to attain their objectives (to win the match), but also allows illicit gamblers, involved

---

[7]Zitzewitz (2012) p. 732.
[8]Zitzewitz (2012) p. 732.

in the fixing, the opportunity to profit. Empirical evidence suggests a clustering of results where favourites win by just less than the predicted spread, but there is some debate as to whether this is due to deliberate point-shaving or simply an unexpected (from the point of view of the bookie) tailing off in effort from the better team as they close in on the win. In a similar vein to ours, Reade and Akie (2013) use football betting data to look for signs of match-fixing. Their idea is that bookmakers have good incentives to sniff out potential fixes, as they will lose money if on the wrong side of the bets. Therefore, match-fixing can potentially be identified by comparing the bookies' odds to the odds produced by prediction models based on public information. When these odds differ substantially, it may be because bookies have spotted a potential fix. Our approach is similar. We expect those providing liquidity (quoting back and lay odds) on the exchange to have a good idea of the level of adverse selection in the market. If they underestimate the level of informed trading, they will lose money to the fixers. If they overestimate the level of informed trading, they will quote uncompetitive prices and be unable to secure trade. Therefore, we can infer the frequency of fixing - in both back and lay bet terms - from the prices that are quoted and executed.

The rest of the paper is structured as follows. In Section 2 we present the model and generate predictions for the pricing of back and lay bets. In Section 3 we describe the data and compare the actual pricing of bets to those predicted by the model. Section 4 concludes.

2   The Model

In this section we present a Glosten and Milgrom-type market microstructure model of betting on an exchange.[9] The central result of these models, when applied to any asset market, is that bid/ask prices deviate from the fundamental value of an asset when there is asymmetric information. The market-maker, whether providing liquidity contractually or voluntarily, will use the bid-ask spread to offset losses to informed traders with gains from uninformed noise traders. The magnitude of the bid-ask spread is increasing in the proportion of informed traders in the population.

---

[9]Easley and O'Hara (1987) extend the Glosten and Milgrom model to account for heterogeneous trade sizes, and Diamond and Verrecchia (1987) include short-selling constraints.

A betting exchange, such as Betfair in the U.K., provides a public limit order book for bets on sporting and related events. Bettors are allowed to take both a long or short position on a particular outcome; consequently those providing liquidity will quote back odds (where the opposing bettor places a stake and receives the odds multiplied by the stake if the outcome occurs) and lay odds (where the opposing bettor receives the stake but is liable for the odds multiplied by the stake if the outcome occurs).

Consider the following model of exchange betting. A market-maker trades a bet on a horse to win a race, where the bet has a fundamental value of either 0 (if the horse loses) or 1 (if the horse wins). This bet is traded with a population of bettors, of whom a proportion μ_B are informed about the winner of the race, and a proportion μ_L are informed that a given horse will lose. The informed bet in a manner consistent with their private information, i.e. backing a horse when it will win, and laying it when it will lose.[10] Bettors can be informed due to the collection of private information (i.e. by studying a horse's form in detail) or by the fixing of races. As noted in the introduction, fixing the winner of a race may require the co-operation of a number of competitors; losing a race does not. This means that $\mu_B \leq \mu_L$. The remainder of the betting population are uninformed about the outcome of the race. Consequently they randomise their bets, betting on the horse to win, with a back bet, with probability 0.5, and betting on the horse to lose, with a lay bet, with probability 0.5.[11]

Prior to any bet, the market-maker assesses the probability of the horse winning as $\lambda \in [0, 1]$. We assume that this prior approximates the empirical probability that the horse wins (i.e. the market-maker has unbiased beliefs). In each instant, a bettor is randomly drawn to bet one unit with the market-maker. The posterior probability that the market-maker assigns to the horse winning is determined by her own prior assessment, the proportion of bettors that are informed that a horse is going to win, and the proportion that are informed

---

[10]We assume that the proportion of informed traders is exogenous. For a model where the decision to fix a sporting event and betting market prices are determined endogenously, see Bag and Saha (2011).

[11]We have assumed - as is typical with these type of microstructure models - that informed traders consume liquidity via market orders, rather than provide liquidity via limit orders. The results of the model hold as long as informed traders predominantly use market orders. As the informed want to ensure their trade is executed - and such a guarantee can only be made with a market order - this should be a reasonable assumption.

that a horse will lose.

Applying Bayes' rule, and simplifying, the market-maker's posterior probability of the horse winning, after a back bet, is

$$\frac{\lambda(1 + \mu_B)}{1 + \mu_B(2\lambda - 1)} \tag{1}$$

and the market-maker's posterior probability of the horse winning, after a lay bet, is

$$\frac{\lambda(1 - \mu_L)}{1 + \mu_L(1 - 2\lambda)} \tag{2}$$

Both back and lay odds are set so as to be regret-free, which means that they incorporate the information contained in the direction of the last order. Therefore, converting the posterior probabilities into odds, we have back odds of

$$B = 1/(\frac{\lambda(1 + \mu_B)}{1 + \mu_B(2\lambda - 1)}) - 1 \tag{3}$$

and lay odds of

$$L = 1/(\frac{\lambda(1 - \mu_L)}{1 + \mu_L(1 - 2\lambda)}) - 1 \tag{4}$$

In Figure 1 we plot the implications of this model for the relationship between implied win probabilities - inferred from the betting odds, and measured in Equations (1) and (2) - and actual win probabilities. For this example, we assume that $\mu_B = \mu_L = 0.2$. In other words, fixing a horse to win is equally as likely as fixing a horse to lose, and 20% of the betting population are informed of the fix. In both panels of Figure 1 we plot implied win probability against actual win probability. Implied win probability is then re-created on the y-axis in orange. This implied win probability is then compared to actual win probabilities, displayed in the form of a smooth green scatterplot. For back bets, horses win less frequently than the odds suggest. This is because the market-maker will be picked off by the informed fixers 20% of the time, and therefore must recoup these losses from the uninformed traders as they randomise their bets. For lay bets there is a symmetrical and opposite relationship: the odds underestimate the frequency with which the horse actually wins. In both cases, the largest difference between implied and actual win probabilities is for bets of around evens (or implied win probabilities near 0.5). This is a characteristic of Bayesian updating. For low or high implied win probabilities the prior is strong, and therefore less is learned from the

7

direction of bettor's trade. Around 0.5, however, the prior is weak, and therefore the effect on the market-maker's beliefs of a given trade is greater.

By way of comparison, in Figure 2 we plot the relationship between implied and actual win probabilities when there is no adverse selection (i.e. $\mu_B = \mu_L = 0$). In this example horse racing is clean and there is no fixing of horses to win or lose. For both back and lay bets, the actual win probabilities closely correspond with the implied win probabilities inferred from the betting odds. The market-maker has no need to recoup losses from uninformed bettors as there are no informed fixers to lose to. Drawing on the patterns in Figures 1 and 2, we can design empirical tests to establish whether betting exchanges, by facilitating lay betting, have corrupted horse racing. If the fixing of horses to lose is now prevalent, we should expect that actual win probabilities will depart from implied win probabilities to a greater extent for lay bets than for back bets. If there is widespread fixing of horses to lose, a regression of actual win probabilities on implied win probabilities, for lay bets, will have a significantly positive intercept coefficient (as illustrated in the bottom panel of Figure 1), as horses win much more frequently than their odds imply. It is this idea that we will test in Section 3.

Before we proceed, it is worth clarifying one point about the model. This is a model of informed trading on the exchange. Bettors can be informed because they have fixed the race, or because they have simply acquired private information. Therefore, if we observe that lay bets have much greater adverse selection inherent in their prices, this is merely consistent with, rather than proof of, the widespread fixing of horses to lose. Lay bets may also have a greater adverse selection component because it is easier to acquire information that a horse is poor and will lose, rather than acquire information that a horse will definitely outperform the competition and win. If, as may be the case, we observe that there is little or no adverse selection associated with lay bets - consistent with our simulated prices in Figure 2 - then this would suggest that betting exchanges have not led to the corruption of horse racing.

3 Data

We collected data on 9,560 U.K. horse races run between the 24th of March 2013 and the 19th March 2014 from Fracsoft, a third-party provider of Betfair limit order book data. In

contrast to previous studies which have just used transaction data (e.g. Smith *et al.* (2006)), we require samples of the limit order book because we need to distinguish between back and lay initiated orders. The data includes second-by-second observations of the best back and lay quotes (odds), plus the last transacted price and volume each second. This data is collected for 20 minutes prior to each race for both the win market and the place market (though certain place markets are missing from the Fracsoft database). In the place market, punters speculate on which horses will finish in top-ranked positions (normally 2nd to 4th, but depending on the number of runners), but will not win the race. This data is also used in a contemporaneous paper (Brown and Yang 2014) which looks, more generally, at the role of speculative trade in asset mispricing, and studies both the pre-race and in-running trading periods.

To classify back and lay initiated orders we use a variation of the Lee and Ready (1991) algorithm. An order occurs if the total matched volume for a bet at time t is greater than the total matched volume at time $t-1$. We then classify the order as a back order, betting on the horse to win/place, if the last matched price at time t is closer to the best quoted back odds at time $t-1$ than it is to the best quoted lay odds at time $t-1$. If the last matched price is closer to the last quoted lay price, then we classify the order as a lay order, betting on the horse not to win/place. This algorithm can classify all orders except those that occur precisely at the middle of the spread at time $t-1$.

Using this method we turn approximately 225 million observations of the limit order book - across the 9,560 races and for both win and place markets - into detailed records of all bets placed. For each bet, we then sum the back and lay orders, and calculate the total back and lay volume and the weighted average back and lay price (described in implied win probability terms where implied win probability=$1/(Odds+1)$). Table 1 summarises this information. From the top panel of this table, we can see that back initiated bets constitute, on average, greater volume than lay initiated bets. The average trading volume is more than 18,000 GBP for back bets, and just over 13,000 GBP for lay bets. The highest trading volume in this 20 minute pre-race window is 1.76 million GBP for back bets, and 2.4 million GBP for lay bets. This is for one bet, on one horse, in one market. We can also see that there is a spread in implied win probabilities between back and lay bets. Back bets have average implied win

probabilities of 0.21 (just less than 4-1 in odds), and lay bets average 0.2. This spread is in part mechanical, as without a spread, back bet quotes would immediately execute with lay bet quotes. Our job later will be to establish the extent to which this spread is due to greater adverse selection in lay betting. Finally, in the middle and bottom panels we break down the summary statistics into the win and place markets. The win market hosts greater betting volume for both back and lay betting.

We marry the Fracsoft data with data from Betfair on the winners and placers in each race.[12] Betfair also provide data on the amount of traded volume and the average price for each bet. We do not use this data because we need to distinguish between back and lay initiated orders. However, by comparing our own measure of trading volume constructed from the Fracsoft data, with the Betfair in-house measure, we estimate that at least 80.4% of pre-race trading takes place in the 20 minutes immediately preceding each race. Moreover, we find a correlation of 0.98 between the total amount of volume that we detect for a bet (both back and lay initiated) and the pre-race volume for that bet listed by Betfair. These two points give us confidence that the prices we describe are a good representation of the prices traded on Betfair.

Our first look at whether betting exchanges have corrupted horse racing can be found in Figure 3. We plot the implied win probability - as inferred from back and lay betting odds - against actual win probabilities (calculated as the average of the win indicator). We round implied win probability to 2 decimal places to ensure that we have large enough buckets of bets to calculate smooth actual win probabilities. We also pool win and place bets to ensure a large spread of implied win probabilities. Judging by the green scatters in Figure 3, it would appear that implied win probabilities calculated from both back and lay bet prices closely shadow actual win probabilities. In other words, bets win/lose about as often as the odds imply. On the face of it this supports a parameterization similar to Figure 2, where there are no informed fixers for back bets or lay bets. In other words, there is little to suggest at this stage that betting exchanges, by allowing for lay betting, have encouraged those competing to deliberately lose. The only area where there is some discrepancy between actual and implied win probabilities is for the extreme favourites (i.e. those with very high implied win/place

---

[12]This can be found here: http://www.betfairpromo.com/betfairsp/prices/index.php.

probabilities). These horses seem to win more often than the odds suggest, both for back and lay bets. Having said that, part of this is due to the limited number of observations up in this top right corner. In general there is more noise here than there is for lower implied win/place probabilities where there are many more observations.

Another way to check the relationship between actual and implied win probabilities is with a lowess (locally weighted scatterplot smoothing) technique. This works by running regressions of the win indicator variable on implied win probabilities for different subsets of the data at a time. In Figure 4 we display the results of this technique for both back and lay bets. For these examples, a bandwidth of 0.8 is chosen. This means that 80% of the data is used for each local regression. This is the default option in many statistical packages, but similar results are found when we vary the bandwidth. Using this technique we find that actual and implied win probabilities are indistinguishable from each other for both back and lay bets. In other words, there is, again, little to suggest that there is an adverse selection component to either back or lay bet prices.

To more formally test whether there is greater adverse selection associated with lay bets, we estimate the following simple regression:

$$y_i = \beta_0 + \beta_1 x_i + \epsilon_i \tag{5}$$

y_i is an indicator variable, equalling 1 if the horse won (or placed, depending on the market), x_i is the implied win probability inferred from the odds, and ϵ_i is an error term. This linear probability model, estimated by OLS, is sometimes used to test for the favourite-longshot bias, where returns for betting on favourites exceed those for betting on longshots (see Reade (2014) for example). In our study - where our later results show no favourite-longshot bias - we are primarily interested in the intercept coefficient β_0. In the presence of adverse selection, the intercept should be negative for back bets, and positive for lay bets. Figure 1, based on our model, illustrates this point. Moreover, if adverse selection is more pronounced for lay bets, the absolute magnitude of this coefficient should be greater for lay bets than it is for back bets. This is because market-makers will revise their beliefs to a greater degree for lay bets, as these types of orders are more likely to stem from informed

trading (after a horse is fixed to lose).[13]

In Table 2 we present the results of this regression for the full sample, and also broken down for the win and place markets. Standard errors are heteroskedasticity-consistent.[14] On the face of it, the model appears to fit the data well. The estimated coefficient for β_1 is insignificantly different from 1, showing an approximate 1:1 relationship between actual and implied win probabilities. This is the case for both back and lay bets, and for both win and place markets. More importantly, we find that the intercept term is insignificantly different from 0 (at least at the 5% level) for both back and lay bets. This suggests that there is no adverse selection component to either back or lay prices. Moreover, for the pooled win and place bets (in the top panel), the absolute magnitude of the intercept is smaller for lay bets (.0002082) than it is for back bets (-.0018182). Therefore, if anything, there is arguably more adverse selection for back bets than there is for lay bets. This strongly suggests that betting exchanges have not become a hive for the fixing of horses to lose.

One possibility is that the fixing of horses to lose is concentrated in low-grade races where the incentives for winning are smaller. This concern was recently raised by the Timeform organisation, a horse racing information company: 'the rules requiring horses to be ridden on their merits are regularly flouted...At times nowadays, at some of the more far-flung outposts of jump racing, it seems as if [stewards] are hardly applying [the rules] at all.'[15] While we cannot identify low grade racing from our data-set, we can identify those races which, as a whole, attract little betting volume. These will tend to be lower grade races at such far-flung outposts. In Table 3, we break down races into those with high and low betting volume. We calculate the total back and lay volume for each race, and a race is classified as high volume if it has volume greater than the median race in the data, and is classified as low volume

---

[13]We could alternatively use a logit specification, to recognise the 0/1 nature of our dependent variable. Or, we could use a conditional logit, to additionally recognise the correlation between the dependent variable within each race. The problem with these models is that our coefficient of interest is the intercept, which would be negative for both back and lay bets in a logit model, and absent in a conditional logit model.

[14]We have observations for each horse, in each market, in each race. This raises the possibility of correlated residuals within races. Similar results to those displayed in the paper - where standard errors are heteroskedasticity-consistent - are found when standard errors are also clustered at the race-level.

[15]'British Horseracing Authority stewards accused of missing non-triers', Chris Cook, *The Guardian*, 1st February 2012.

otherwise. This process is carried out separately for win and place markets, as volume differs substantially across the two markets. We then carry out the same regression described in Equation (5) for these sub-samples. We find little evidence consistent with lay bet fixing in low volume races. The intercept coefficient is insignificant for lay bets in low volume races, in both the win and place markets. Actual win probabilities closely correspond with implied win probabilities, suggesting that bettors do not pay a premium consistent with widespread fixing of horses to lose.

One other possibility is that adverse selection for lay bets is only detectable when a large, disproportionate, amount of money is placed on a horse to lose. A group of individuals are only likely to risk the punishment of cheating if they can profit handsomely; to do so they must bet a large amount. With this in mind, we calculated the net volume - which is the total back volume minus the total lay volume - on each horse, in each market (win or place), in each race. While some horses may attract a large number of lay bets, this may be offset by a large number of back bets. We therefore want to know what is an abnormally high level of lay betting, which depends not only the amount bet both for and against that horse, but also on the amount bet for and against other horses in the same race. Once we had calculated the net volume for each horse, we then identified those horses in the bottom 10% and 25% of net volume in their race. If abnormally large amounts of lay volume do correlate with the fixing of horses to lose, we would expect these groups of horses to win more often than their odds suggest.[16] To this end, we repeat the analysis of Table 2, but this time with our sub-sample of horses with abnormally large laying volume. These new results are in Table 4. In line with our earlier results, we find an insignificant intercept coefficient for lay bets, and the fact that this coefficient is negative indicates that these horses, if anything, actually win slightly less often than their odds suggest. This is the case both for the bottom 10% of net volume within each race (in the top panel), and the bottom 25% of net volume (in the bottom panel). In short, a look at extreme lay volume does not turn up any evidence of corruption.

Our next piece of analysis focuses on the number of runners within a race. The ease

---

[16]It is possible, of course, to end up in these sub-samples purely by attracting fewer back bets than others in the same race, rather than attracting more lay bets; however, this is not the case for the majority of horses in these sub-samples.

with which the winner of a race can be fixed likely depends on the number of runners. The more runners there are, the less easy it is to fix, as exercising control over more and more runners becomes impractical. Fixing a horse to lose, however, is less sensitive to the number of runners, as the fixer only has to control one horse. Similarly, suppose a punter has fixed a horse to lose, but wants to hide their activity so instead of laying the horse they back all of the horses' competitors. When the number of runners becomes large, however, this becomes costly (transaction costs are paid on each bet), and therefore lay-fixing in large fields (with many runners) will likely be enacted exclusively through lay bets not back bets. Following these two arguments, it is possible that the degree of adverse selection for lay bets, relative to adverse selection for back bets, is more pronounced for large races. With this in mind, we broke the sample approximately in half, and looked at races with less than 10 runners, and races with 10 or more runners. In Table 5 we present the results of this analysis. If lay-fixing is widespread, for large races we should observe a greater difference between the implied and actual win probabilities for lay bets than for back bets. On the contrary, for races with 10 or more runners, we find that the intercept term is insignificant for both back and lay bets, and that it is, again, the back bet prices which depart to a greater extent from the empirical probabilities. This indicates, if anything, that there is a sliver more adverse selection in back bets than in lay bets, even in races where it is hardest to fix the winner.

Our final tests break down the full sample into implied win probability sub-samples. From Figure 3 there appeared to be some underestimation of the win probabilities of extreme favourites. Much of this appeared to be due to the limited number of observations, but it is worth checking if this underestimation is statistically significant. With this in mind, in Table 6 we consider four separate sub-samples of implied win probability. We can see from the number of observations that there are many more bets with implied win/place probabilities less than 0.25 (112,186 for back bets and 113,580 for lay bets) than there are for implied win probabilities greater than 0.75 (2697 for back bets and 2583 for lay bets). The intercept terms for these lay bets is large (.1015566) but is not statistically significant, giving weight to the explanation that discrepancies between actual and implied win probabilities in Figure 3 were due to the small number of observations. Moreover, the intercept term is actually bigger for back bets (.1165823) when, if adverse selection were the cause, we would expect

back bets to have a negative intercept coefficient. In short, our analysis has turned up no evidence that adverse selection is pronounced for lay bets, and therefore no evidence that betting exchanges have brought about the corruption of horse racing.

Before we conclude, we will discuss a number of caveats with our analysis. Firstly, it is possible that even substantial levels of corruption may be difficult to detect in the data. For example, suppose that 1 in 10 races are afflicted by competitors deliberating losing. Within this 1 race, suppose that this affliction only relates to 1 out of the 10 runners. If this is the case, adverse selection only affects 1 in 100 bets, and such a small effect is therefore likely to be economically and statistically insignificant. However, even if the 'treatment' is tiny - and only affects a small fraction of bets - we might expect there to be a slight indication that adverse selection is more of a threat for lay bets than back bets. Yet, our point estimates for our intercept terms suggest that, if anything, there is greater adverse selection inherent in back bet prices than in lay bet prices. This is the case even in sub-samples - such as for low-grade races, and for horses with abnormally large lay volumes - where we might expect lay-fixing to at least be more pronounced (if still not statistically and economically significant).

A second caveat is that our work relates to only one year of racing (in part due to the number of order book observations that even a small number of races generate). We chose a recent window in order to answer the question of whether horse racing was recently, or currently, corrupted. It is possible, however, that the pricing of bets in earlier years of the betting exchanges may indicate greater levels of adverse selection for lay bets. Our final caveat is that the development of betting exchanges may have necessitated elevated levels of expenditure on race-fixing detection. We are aware of systems that the British Horse Racing Authority have in place, in concert with Betfair, which aim to create a clear audit trail to detect the fixing of horses to lose. It is possible that without such expenditures, the invention of betting exchanges may have had substantial adverse consequences for the integrity of horse racing.

4   Conclusion

Betting exchanges have shaken up the way that sports betting takes place in the U.K.. Rather than relying on the odds offered by bookmakers - with their decades of experience and high-street presence - bettors can now offer bets amongst themselves. There is evidence that this has substantially driven down the cost of betting (Smith *et al.* (2006)). Nevertheless, there is the suspicion that this may have come at a grave cost for the most prominent sport for betting, horse racing. Incidences of horses being fixed to lose - in order that their connections can profit from lay betting on the exchanges - are numerous. This has fed the impression that there is a new widespread form of corruption afflicting horse racing.

Measuring the extent to which anecdotal evidence of corruption is indicative of a more widespread problem is, of course, difficult. Wrongdoers take steps to hide their activities, in order to avoid punishment. Our approach in this paper, therefore, is to build an asymmetric information model of betting exchange activity and then take it to the data. The central idea in our model is that if the deliberate fixing of horses to lose is widespread, then horses subject to lay bets should win much more often than their odds imply. This is because counterparties to any corrupt bettors must recoup their losses from other unsuspecting noise bettors. When we examined a recent year of betting exchange data, however, there was little sign of adverse selection in the pricing of lay bets. Actual win probabilities closely corresponded with implied win probabilities, suggesting that the examples of horses being fixed to lose are not signs of a pervasive problem in the sport.

A note of caution should nevertheless be struck. Horse race betting is big business in the U.K.. In our 1-year sample, bets with a total value exceeding 6 billion GBP were struck. The rewards from fixing horse races are therefore great. It is also undoubtedly true that it is operationally easier to fix a horse to lose a race, than it is to fix a horse to win. We hope that the simple theoretical and empirical framework presented in this paper will help in detecting whether the fixing of races becomes a problem in the future.

In addition, the method we use in this paper may be useful for detecting corruption in other sports, such as golf, cycling and Formula 1. As in each of these sports it is easier to fix a loser than it is to fix a winner, we can examine differences in returns to back and lay bets

to see if betting market-induced corruption has materialised. While racing and individual sports with many competitors are the most obvious application, there is also the potential to apply a similar method in other sports. For example, a comparable method may be useful for determining corruption in so-called 'spot-betting' markets. In cricket, punters can wager on the number and timing of 'no-balls', which are non-legitimate deliveries which lead to a run being awarded to the opposing batting side. There may be differences in the returns to betting on the number of no-balls in a single over - which is in the control of a single bowler - and the number of no-balls in a series of overs, where responsibility is spread across more than one bowler. In much the same way that horse racing fixers will find it easier to fix a horse to lose than to win (as they just need to control one horse), cricket fixers will find it easier to fix the number of no-balls in a single over than in a series of overs (as they just need to control one bowler). This should depress returns for bets on a single over - as other bettors aim to recoup money lost to the fixers - relative to bets on a series of overs.

References

- Bag, P., K., Saha, B., (2011). Match-Fixing under Competitive Odds. *Games and Economic Behavior*, 73, 318-344.

- Bernhardt, D., Heston, S., (2010). Point Shaving in College Basketball: A Cautionary Tale for Forensic Economics. *Economic Inquiry*, 48, 14-25.

- Brown, A., Yang, F., (2014). The Role of Speculative Trade in Market Efficiency: Evidence from a Betting Exchange. *mimeo.*

- Diamond, D., W., Verrecchia, R., E., (1987). Constraints on Short-Selling and Asset Price Adjustment to Private Information. *Journal of Financial Economics*, 18, 277-311.

- DellaVigna, S., Ferrara, E., L., (2010). Detecting Illegal Arms Trade. *American Economic Journal: Economic Policy*, 2, 26-57.

- Easley, D., O'Hara, M., (1987). Price, Trade Size, and Information in Securities Markets. *Journal of Financial Economics*, 19, 69-90.

- Glosten, L., R., Milgrom, P., R., (1985). Bid, Ask and Transaction Prices in a Specialist Market with Heterogeneously Informed Traders. *Journal of Financial Economics*, 14, 71-100.

- Lee, C., M., C., Ready, M., J., (1991). Inferring Trade Direction from Intraday Data. *Journal of Finance*, 46, 733-746.

- O'Connor, N., (2014). *A History of the Betting Exchange Industry.* www.bettingmarket.com

- Reade, J., J., (2014). Information and Predictability: Bookmakers, Prediction Markets and Tipsters as Forecasters. *Journal of Prediction Markets*, forthcoming.

- Reade, J., J., Akie, S., (2013). Using Forecasting to Detect Corruption in International Football. RPF Working Paper No. 2013-005, http://www.gwu.edu/ forcpgm/2013-005.pdf.

- Shin, H., S., (1993). Measuring the Incidence of Insider Trading in a Market for State-Contingent Claims. *The Economic Journal*, 103, 1141-1153.

- Smith, M., A., Paton, D., Vaughan Williams, L., (2006). Market Efficiency in Person-to-Person Betting. *Economica*, 73, 673-689.

- Wolfers, J., (2006). Point Shaving: Corruption in NCCA Basketball. *American Economic Review*, 96, 279-283.

- Zitzewitz, E., (2012). Forensic Economics. *Journal of Economic Literature*, 50, 731-769.
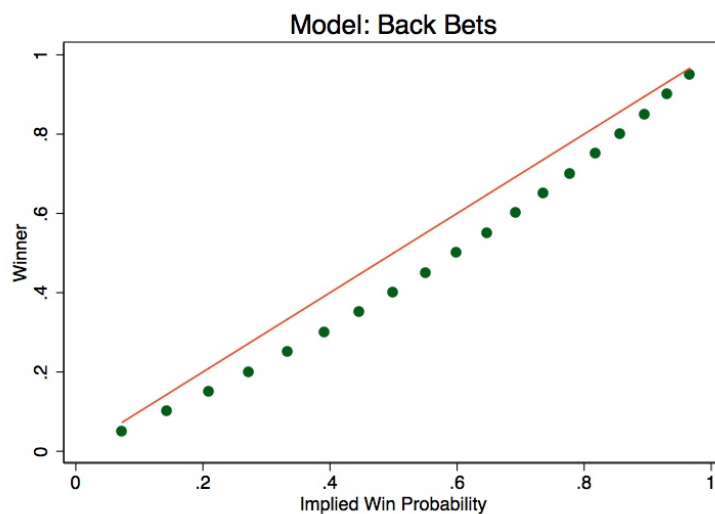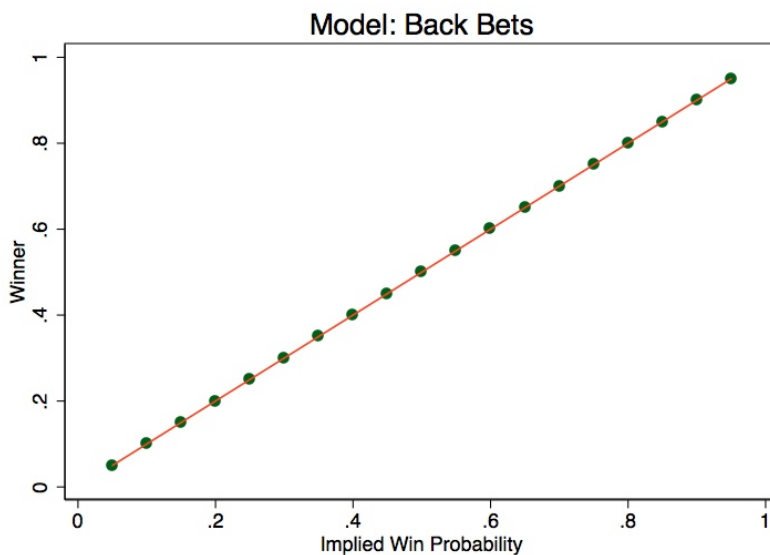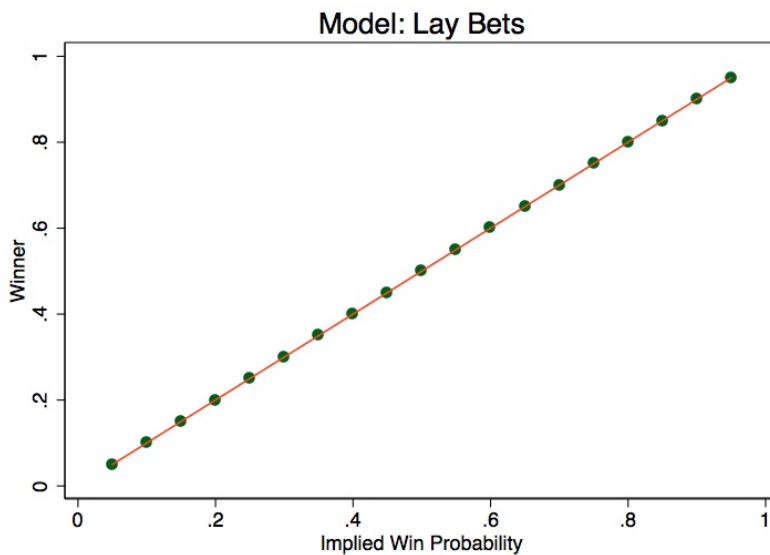
(a) $\mu_B = 0.2$
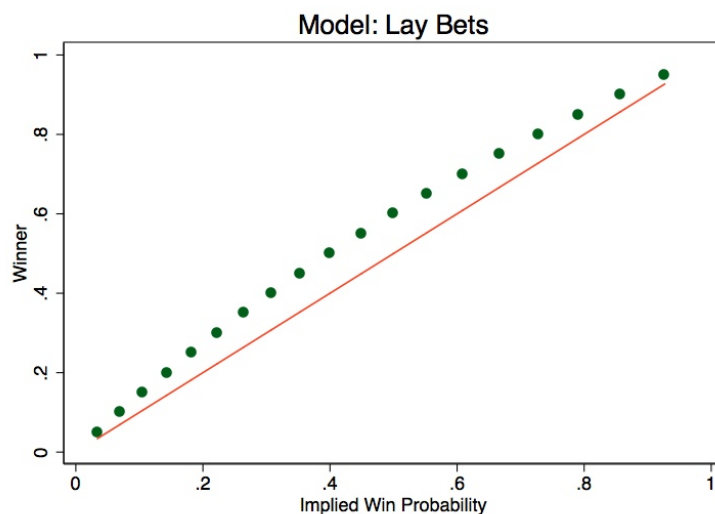
(b) $\mu_L = 0.2$

Figure 1: The model's predicted relationship between the implied win probability of a bet (captured by the orange straight line) and the actual win probability of the bet (captured by the green scatter plot), in the presence of adverse selection. The top panel contains back bets, and the bottom panel contains lay bets. For both types of bet, the proportion of informed bettors in the population is assumed to be 20% (i.e. $\mu_B = \mu_L = 0.2$). If there is adverse selection - as is the case here - the horse subject to a back (lay) bet wins less (more) often than the odds imply.

(a) $\mu_B = 0$

(b) $\mu_L = 0$

Figure 2: The model's predicted relationship between the implied win probability of a bet (captured by the orange straight line) and the actual win probability of the bet (captured by the green scatter plot), with *no* adverse selection. The top panel contains back bets, and the bottom panel contains lay bets. For both types of bet, the proportion of informed bettors in the population is assumed to be 0 (i.e. $\mu_B = \mu_L = 0$). If there is no adverse selection - as is the case here - the horse subject to a back (lay) bet wins equally as often as the odds imply.
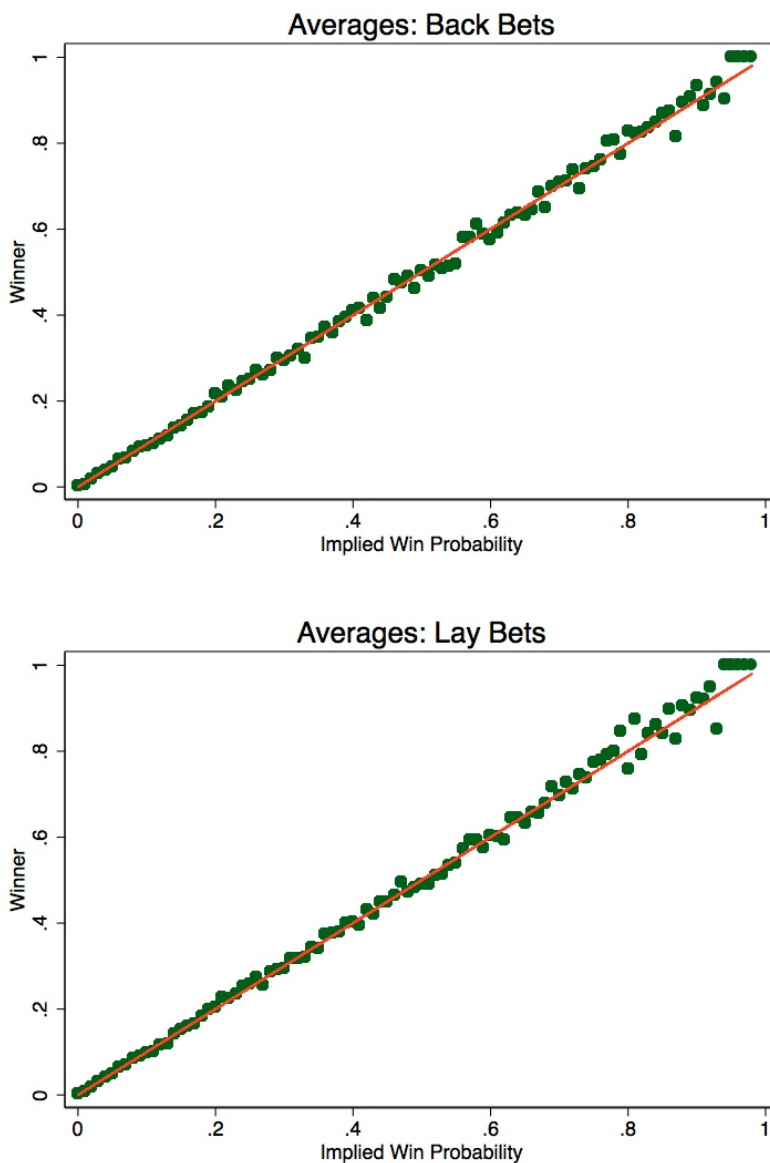
Figure 3: The empirical relationship between the implied win probability of a bet (captured by the orange straight line) and the actual win probability of the bet (captured by the green scatter plot). The top panel contains back bets, and the bottom panel contains lay bets. Implied Win Probability=$1/(Odds + 1)$. Implied win probability is rounded to 2 decimal places, to create large enough buckets to compare with actual win probabilities. For both back and lay bets, the data are consistent with no adverse selection (i.e. $\mu_B = \mu_L = 0$)
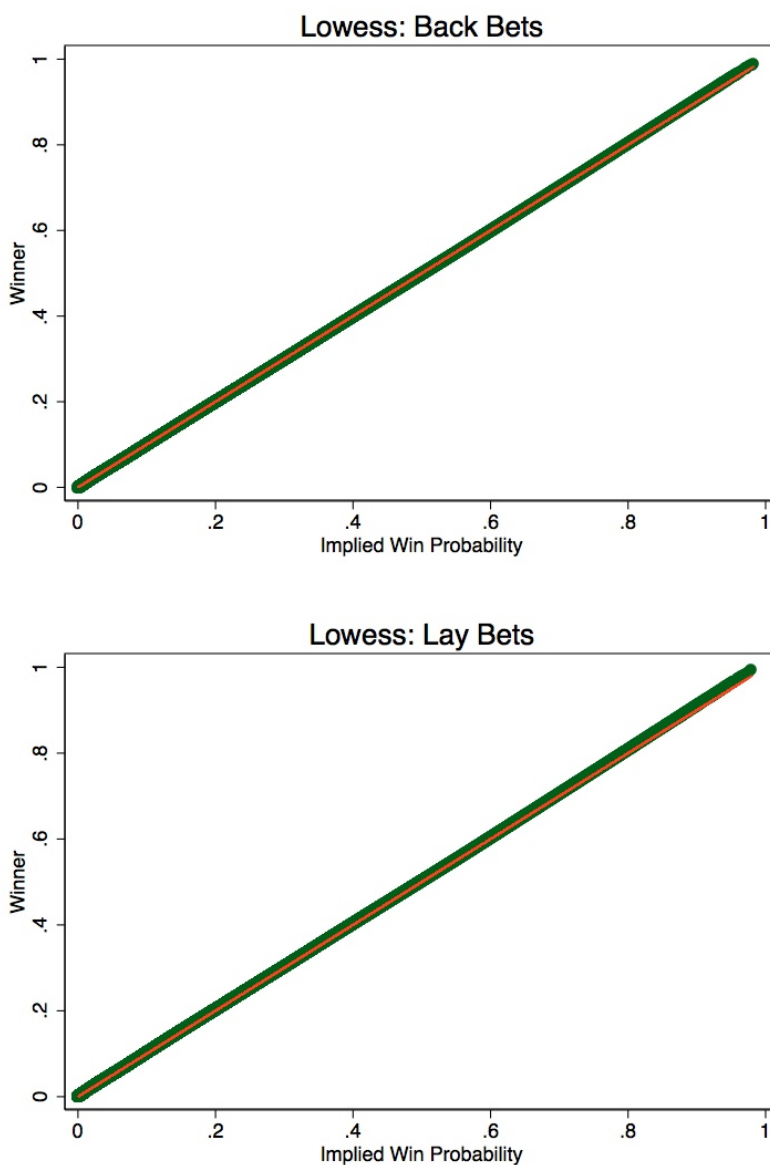
Figure 4: A Lowess of the empirical relationship between the implied win probability of a bet (captured by the orange straight line) and the actual win probability of the bet (captured by the green scatter plot). The top panel contains back bets, and the bottom panel contains lay bets. Implied Win Probability$=1/(Odds+1)$. A bandwidth of 0.8 is used for the Lowess. For both back and lay bets, the data are consistent with no adverse selection (i.e. $\mu_b = \mu_l = 0$).

Table 1: Summary Statistics					
All (Win & Place Markets)					
Total Trading Volume	Observations	Mean	Std. Dev.	Min.	Max.
Back-initiated Bets	169813	18.36734	40.43982	.002	1759.004
Lay-initiated Bets	170141	13.13567	31.51797	.002	2395.698
Implied Win Probability	Observations	Mean	Std. Dev.	Min.	Max.
Back-initiated Bets	169813	.2133752	.1902648	.0010101	.9826381
Lay-initiated Bets	170141	.2099256	.1893823	.001	.9799851
Win Market					
Total Trading Volume	Observations	Mean	Std. Dev.	Min.	Max.
Back-initiated Bets	85670	31.2845	53.52958	.002	1759.004
Lay-initiated Bets	86058	22.21862	42.0831	.004	2395.698
Implied Win Probability	Observations	Mean	Std. Dev.	Min.	Max.
Back-initiated Bets	85670	.1158333	.1162061	.0010101	.9752696
Lay-initiated Bets	86058	.1134978	.1156807	.001	.9615385
Place Market					
Total Trading Volume	Observations	Mean	Std. Dev.	Min.	Max.
Back-initiated Bets	84143	5.215774	6.339611	.00246	157.1577
Lay-initiated Bets	84083	3.839365	5.163646	.002	89.16811
Implied Win Probability	Observations	Mean	Std. Dev.	Min.	Max.
Back-initiated Bets	84143	.3126873	.1993981	.0016186	.9826381
Lay-initiated Bets	84083	.3086184	.1990493	.0010101	.9799851

Summary statistics on Betfair betting on the runners in 9,560 horse races in the U.K. between the 24th March 2013 and the 19th March 2014. Win and place bets for each horse are used, and back/lay -initiated bets are categorised by a variation of the Lee and Ready (1991) algorithm described in Section 3. Total trading volume is measured in thousands of GBP, and Implied Win Probability=$1/(Odds+1)$. Win and place market bets are pooled in the top panel, and separated out in the middle and bottom panels.

Table 2: Main Regression Analysis

Dependent Variable: Win Indicator

All (Win & Place Markets)

	Coefficient	Std. Error	t-statistic	p-value	95% CI Lower	95% CI Upper	R^2	No. of Obs
Back bets								
Intercept	-.0018182	.0010217	-1.78	0.075	-.0038207	.0001844	0.2168	168052
Implied Win Probability	.9993004	.00503	198.67	0.000	.9894417	1.009159		
Lay bets	Coefficient	Std. Error	t-statistic	p-value	95% CI Lower	95% CI Upper	R^2	No. of Obs
Intercept	.0002082	.0010131	0.21	0.837	-.0017775	.0021939	0.2171	168384
Implied Win Probability	1.003939	.0050542	198.63	0.000	.9940331	1.013845		

Win Market

	Coefficient	Std. Error	t-statistic	p-value	95% CI Lower	95% CI Upper	R^2	No. of Obs
Back bets								
Intercept	-.0010515	.001254	-0.84	0.402	-.0035094	.0014063	0.129	83942
Implied Win Probability	.9982573	.0123883	80.58	0.000	.9739763	1.022538		
Lay bets	Coefficient	Std. Error	t-statistic	p-value	95% CI Lower	95% CI Upper	R^2	No. of Obs
Intercept	.0002896	.0012344	0.23	0.815	-.0021299	.002709	0.1296	84334
Implied Win Probability	1.002004	.0124223	80.66	0.000	.9776568	1.026352		

Place Market

	Coefficient	Std. Error	t-statistic	p-value	95% CI Lower	95% CI Upper	R^2	No. of Obs
Back bets								
Intercept	-.0035778	.0021406	-1.67	0.095	-.0077733	.0006176	0.187	84110
Implied Win Probability	1.002862	.0066523	150.75	0.000	.9898236	1.015901		
Lay bets	Coefficient	Std. Error	t-statistic	p-value	95% CI Lower	95% CI Upper	R^2	No. of Obs
Intercept	.0003573	.0021289	0.17	0.867	-.0038153	.00453	0.1866	84050
Implied Win Probability	1.003897	.0066793	150.30	0.000	.9908055	1.016988		

Regression results for the full sample. An indicator equalling 1 if the horse won (or placed, depending on the market) is regressed on the implied win probability inferred from the betting odds. Implied Win Probability=$1/(Odds+1)$. Back and lay bets are considered in separate regressions. The top panel displays the results for pooled regressions of win and place market bets, while the win and place markets are considered separately in the middle and bottom panels. Standard errors are heteroskedasticity-consistent.

Table 3: High/Low Volume Races

Dependent Variable: Win Indicator

	Coefficient	Std. Error	t-statistic	p-value	95% CI Lower	95% CI Upper	R^2	No. of Obs
High Volume in Win Market								
Back bets							R^2	No. of Obs
Intercept	.0006235	.0016967	0.37	0.713	-.0027021	.003949	0.1588	39947
Implied Win Probability	.9862902	.0156687	62.95	0.000	.9555792	1.017001		
Lay bets							R^2	No. of Obs
Intercept	.0018762	.0016695	1.12	0.261	-.001396	.0051484	0.1594	40188
Implied Win Probability	.98943	.0156964	63.04	0.000	.9586648	1.020195		
Low Volume in Win Market								
Back bets							R^2	No. of Obs
Intercept	-.00349	.0018746	-1.86	0.063	-.0071642	.000184	0.0976	43995
Implied Win Probability	1.019307	.0200608	50.81	0.000	.9799877	1.058627		
Lay bets							R^2	No. of Obs
Intercept	-.0021292	.0018457	-1.15	0.249	-.0057468	.0014884	0.0979	44146
Implied Win Probability	1.024461	.0201476	50.85	0.000	.9849713	1.063951		
High Volume in Place Market								
Back bets							R^2	No. of Obs
Intercept	-.003789	.0027648	-1.37	0.171	-.0092079	.00163	0.2029	44902
Implied Win Probability	1.003048	.0086794	115.57	0.000	.986036	1.020059		
Lay bets							R^2	No. of Obs
Intercept	-.0000716	.0027486	-0.03	0.979	-.0054589	.0053158	0.2026	44887
Implied Win Probability	1.004346	.0087126	115.28	0.000	.9872696	1.021423		
Low Volume in Place Market								
Back bets							R^2	No. of Obs
Intercept	-.0033	.0033773	-0.98	0.329	-.0099196	.0033196	0.1687	39208
Implied Win Probability	1.002546	.0103665	96.71	0.000	.9822277	1.022865		
Lay bets							R^2	No. of Obs
Intercept	.0009236	.0033606	0.27	0.783	-.0056632	.0075104	0.1683	39163
Implied Win Probability	1.003167	.0104119	96.35	0.000	.9827598	1.023575		

Regression results for high and low betting volume races. An indicator equalling 1 if the horse won (or placed, depending on the market) is regressed on the implied win probability inferred from the betting odds. Implied Win Probability=$1/(Odds+1)$. Back and lay bets are considered in separate regressions, as are win and place markets.

25

Table 4: Large Lay Volume

Dependent Variable: Win Indicator

Lowest Net Volume within Race (Bottom 10%)

	Coefficient	Std. Error	t-statistic	p-value	95% CI Lower	95% CI Upper	R^2	No. of Obs
Back bets							0.2819	7192
Intercept	-.0022095	.0033631	-0.66	0.511	-.0088021	.004383		
Implied Win Probability	1.022355	.0237192	43.10	0.000	.975858	1.068851		
Lay bets							0.2835	7314
Intercept	-.0004215	.0032748	-0.13	0.898	-.0068411	.0059981		
Implied Win Probability	1.028308	.0238364	43.14	0.000	.9815817	1.075034		

Lowest Net Volume within Race (Bottom 25%)

	Coefficient	Std. Error	t-statistic	p-value	95% CI Lower	95% CI Upper	R^2	No. of Obs
Back bets							0.2813	34746
Intercept	-.0025688	.0015716	-1.63	0.102	-.0056493	.0005117		
Implied Win Probability	1.000069	.0108941	91.80	0.000	.9787159	1.021421		
Lay bets							0.2827	35080
Intercept	-.0009791	.0015441	-0.63	0.526	-.0040056	.0020474		
Implied Win Probability	1.006775	.01095	91.94	0.000	.9853125	1.028237		

Regression results for horses that received the lowest net volume (i.e. the greatest lay volume) in their race. An indicator equalling 1 if the horse won (or placed, depending on the market) is regressed on the implied win probability inferred from the betting odds. Implied Win Probability=$1/(Odds+1)$. Back and lay bets are considered in separate regressions. The top panel displays the 10% of horses with the lowest net volume in their race, while the bottom panel displays the 25% of horses with the lowest net volume. Standard errors are heteroskedasticity-consistent.

Table 5: No. of Runners

Dependent Variable: Win Indicator

	Coefficient	Std. Error	t-statistic	p-value	95% CI Lower	95% CI Upper	R^2	No. of Obs
Runners<10								
Back bets							0.228	79994
Intercept	-.0021079	.0016757	-1.26	0.208	-.0053924	.0011765		
Implied Win Probability	1.001815	.0064824	154.54	0.000	.98911	1.014521		
Lay bets							0.2281	80065
Intercept	.0000573	.0016659	0.03	0.973	-.0032079	.0033225		
Implied Win Probability	1.005779	.0065118	154.46	0.000	.9930162	1.018542		
Runners ≥ 10								
Back bets							0.1814	88058
Intercept	-.0012912	.0013145	-0.98	0.326	-.0038676	.0012852		
Implied Win Probability	.9942807	.0085071	116.88	0.000	.9776069	1.010954		
Lay bets							0.1818	88319
Intercept	.0005636	.0012998	0.43	0.665	-.0019841	.0031113		
Implied Win Probability	1.000079	.0085584	116.85	0.000	.9833046	1.016853		

Regression results for races broken down into sub-samples based on the number of runners (horses). An indicator equalling 1 if the horse won (or placed, depending on the market) is regressed on the implied win probability inferred from the betting odds. Implied Win Probability=$1/(Odds+1)$. Back and lay bets are considered in separate regressions. The top panel displays results for races with less than 10 runners, with races with 10 or more runners in the bottom panel. Standard errors are heteroskedasticity-consistent.

Table 6: Favourites and Longshots

Dependent Variable: Win Indicator

	Coefficient	Std. Error	t-statistic	p-value	95% CI Lower	95% CI Upper	R^2	No. of Obs
Implied Win Probability<0.25								
Back bets							0.0559	112186
Intercept	-.0025831	.0011429	-2.26	0.024	-.0048231	-.000343		
Implied Win Probability	1.009298	.0137518	73.39	0.000	.9823449	1.036252		
Lay bets							0.0574	113580
Intercept	-.0016562	.0011245	-1.47	0.141	-.0038602	.0005478		
Implied Win Probability	1.026124	.0137337	74.72	0.000	.9992065	1.053042		
0.25<=Implied Win Probability<0.5								
Back bets							0.0215	39191
Intercept	.0017254	.0120962	0.14	0.887	-.0219835	.0254343		
Implied Win Probability	.989226	.0339112	29.17	0.000	.9227592	1.055693		
Lay bets							0.0214	38519
Intercept	.0049519	.0122653	0.40	0.686	-.0190885	.0289923		
Implied Win Probability	.990614	.0343297	28.86	0.000	.923327	1.057901		
0.5<=Implied Win Probability<0.75								
Back bets							0.0206	13978
Intercept	-.0239035	.0355427	-0.67	0.501	-.0935719	.0457649		
Implied Win Probability	1.02832	.0587113	17.51	0.000	.9132384	1.143402		
Lay bets							0.0217	13702
Intercept	-.0382271	.0358897	-1.07	0.287	-.1085758	.0321217		
Implied Win Probability	1.059289	.0592361	17.88	0.000	.9431777	1.175399		
0.75<=Implied Win Probability<1								
Back bets							0.0174	2697
Intercept	.1165823	.0949033	1.23	0.219	-.0695084	.302673		
Implied Win Probability	.8738857	.1120992	7.80	0.000	.6540765	1.093695		
Lay bets							0.0175	2583
Intercept	.1015566	.0988381	1.03	0.304	-.0922534	.2953665		
Implied Win Probability	.8964895	.1171517	7.65	0.000	.6667687	1.12621		

Regression results for favourites and longshots. An indicator equalling 1 if the horse won (or placed, depending on the market) is regressed on the implied win probability inferred from the betting odds. Implied Win Probability=$1/(Odds+1)$. Back and lay bets are considered in separate regressions, and win and place bets are pooled.