

# The Role of Speculative Trade in Market Efficiency: Evidence from a Betting Exchange\*

Alasdair Brown<sup>†</sup>

Fuyu Yang<sup>‡</sup>

University of East Anglia

University of East Anglia

April 19, 2016

## Abstract

Does speculative trade reduce mispricing and help create efficient markets or does it drive prices further from fundamentals? We analyse betting exchange trading on 9,562 U.K. horse races in 2013 and 2014 to find out. Crucially, as each race is run, the fundamental value of bets is unambiguously revealed. We find that the volume of trade is predictive of fundamentals, suggesting that speculative trade is on average conducive to market efficiency. However, much of this effect is concentrated in the in-running period during races when, even without trade, asset fundamentals would be revealed seconds later.

JEL Classification: G02, G12, G14

Keywords: market efficiency, trading volume, asset fundamentals, betting markets

---

\*We would like to thank Burton Hollifield, the Editor, and an anonymous referee for very detailed comments and suggestions. We are also grateful to Raphael Markellos, Peter Moffatt, Arnold Polanski and an audience at UEA for their feedback. Sue Carpenter provided excellent copy-editing assistance. Alasdair Brown would like to acknowledge financial support for this project from the British Academy and the Leverhulme Trust (Ref: SG140097). The research presented in this paper was carried out on the High Performance Computing Cluster supported by the Research and Specialist Computing Support service at the University of East Anglia.

<sup>†</sup>Corresponding Author. School of Economics, University of East Anglia, Norwich NR4 7TJ, U.K.. Email: [alasdair.brown@uea.ac.uk](mailto:alasdair.brown@uea.ac.uk)

<sup>‡</sup>School of Economics, University of East Anglia, Norwich NR4 7TJ, U.K.. Email: [fuyu.yang@uea.ac.uk](mailto:fuyu.yang@uea.ac.uk)

# 1 Introduction

What role does speculative trade play in market efficiency? Do speculators reduce mispricing by buying up cheap assets and selling those that are expensive? Or do speculators—with their propensity for overconfidence, erroneous beliefs, and herd behaviour—buy and sell the wrong assets, drive prices further from fundamentals, and make markets less efficient? These have always been important questions in finance but have assumed greater importance in light of the heightened trading activity of recent years. Chordia *et al.* (2011) document that share turnover on the New York Stock Exchange increased more than five-fold from 1993 to 2008. The Bank of International Settlements estimates that daily turnover in foreign exchange markets was \$5.3 trillion in 2013, up from \$4 trillion in 2010 (Triennial Central Bank Survey, 2013). If trade has a distorting effect on asset prices, the case for increasing trading costs, perhaps through a transactions tax, becomes stronger.

The problem in establishing the role that speculative trade has on market efficiency is that asset fundamentals are seldom observed. Stocks, to take the most prominent financial asset, are infinitely-lived.<sup>1</sup> One possibility is to look at the immediate effect of trading volume on returns and, if these returns do not subsequently reverse, conclude that trading volume reduces mispricing and increases market efficiency. However there is a mechanical element to such a test—as trading creates price pressure unrelated to information—and any analysis is sensitive to the time-horizon chosen for reversal and is therefore less than satisfactory.

Theory is ambiguous regarding the role of speculative trade in market efficiency. In Milgrom and Stokey (1982), no trade takes place in a speculative market as the willingness to trade signals private information and adverse selection. On the other hand, add liquidity traders who participate in the market for non-informational reasons, as in Glosten and Milgrom (1985), and trade does take place, and in fact drives prices towards fundamental values. The presence of liquidity traders allows the informed to hide.

An alternative view is that noise traders may drive prices further from fundamental values. In

---

<sup>1</sup>Even for financial assets that are finitely-lived, we cannot cleanly observe fundamentals. In option markets, for example, suppose there is heavy buying of an out-of-the-money call option which subsequently finishes in the money. The relationship between trading volume and fundamentals, however, is endogenous as the counterparty to the option trade may have hedged their exposure in the stock market thereby driving up its price and increasing the likelihood that the option finishes in-the-money. This could be unconnected to firm fundamentals.

De Long *et al.* (1990), noise traders governed by sentiment bid up the price of already over-priced assets. The arbitrageurs in their model, rather than correcting this mispricing, sometimes join in as short-term returns are higher for over-valued assets. This ‘positive-feedback trading’ drives prices further from fundamentals and creates larger bubbles. Another channel through which trading activity may be distortionary is when traders exhibit herd behaviour, as in Avery and Zemsky’s (1998) model. In such a model, traders ignore their own private information signal, trade based on inferences from the trades of others and therefore amplify any mispricings.

In short, there is no single clear prediction of whether or not, on average, speculative trade moves prices towards fundamentals and thereby removes mispricing.

In the spirit of Angrist and Pischke (2009), imagine the perfect empirical test. Suppose firms lived for only one day. Firms raised capital in equity markets at the beginning of each day, and these stocks were traded from morning until late afternoon. At the end of the afternoon, each firm was wound up and a liquidating dividend was paid to each of the final stock holders. In such a situation, we could sort all stocks by the trading volume that took place throughout the day. We could then compare the returns for those stocks that received pronounced volume with those that did not. If returns for the former were higher, trade would appear to drive prices toward fundamentals, measured by the size of the liquidating dividend, and thus increase market efficiency. If returns for those with large volume were lower, then on balance speculative trade would appear to have a detrimental impact on market efficiency.

While such a scenario is impossible in equity markets, it closely resembles trading in U.K. horse race betting markets. Bets are traded throughout the day on Betfair, a public limit order book similar in structure to the major financial exchanges. Traders submit market and limit orders, and either bet on (buy) or against (sell) a given horse. The race is then run and the fundamental value of each bet is unambiguously revealed.<sup>2</sup>

In this paper we analyse Betfair trading on 9,562 U.K. races run in 2013 and 2014. We use approximately 225 million observations of the limit order book to extract data on more than 60

---

<sup>2</sup>We are assuming that horse race fundamentals are unaffected by the trading volume prior to and during the race. In other words, we assume that the jockey does not adapt his/her effort after observing a number of favourable/unfavourable bets. As our strongest results relate to the in-running period, when jockeys cannot check betting activity, we do not believe that this is a strong assumption. Note that this does not preclude the fixing of races; it only precludes the decision to fix a race after having seen the betting in the last 20 minutes prior to the race.

million separate trades. We then divide bets up by comparing the actual win probabilities, or win frequencies, of those bets that received trading volume above the median within their race with those that did not. Importantly, we are able to control for the well-known favourite-longshot bias by using nearest neighbour matching methods to pair up horses with very similar implied win probabilities, using the average betting prices. In effect, we are asking whether a relatively large amount of volume is predictive of a horse's fundamental probability of winning? Put another way, if two horses are trading at the same price but one receives a great deal more bets at this price, does this indicate that the high volume horse is more likely to win than his/her odds suggest? If so, this would imply that speculative trade is, on average, conducive to market efficiency as mispriced quotes and valuations attract a disproportionate amount of trade.

We also exploit two dimensions of variation in trading on the exchange. Trading takes place in the win market where the identity of the first-placed horse is speculated upon, but also in the place market where the identities of those that finish in the top 2 places (or the top 3 or 4 places, depending on the number of runners<sup>3</sup>). In some ways the place market can be likened to the stock market and the win market to a corresponding option market. Trade in the place market involves a broad bet that a horse will do well: in much the same way, a buy trade in the stock market speculates that a stock is undervalued. Betting in the win market, however, is a much more precise bet that the horse will do *very* well: in much the same way a trader can buy a heavily out-of-the-money call option to speculate that the underlying stock is grossly undervalued.

Our second source of variation is produced by a clean separation in trading period. Trading takes place in the hours before races and also in the 1-5 minutes in which a race is run. We liken the period prior to racing to trading prior to earnings announcements, when all information that is publicly available is stale. The in-running period therefore corresponds to trading during and immediately after earnings announcements, when traders chew over the implications of the new information and incorporate it into asset prices.

These two sources of variation lead to a mixture of results. We find that the average trade prior to races is not consistently predictive of fundamentals. For example, in our preferred matching specification, a horse that receives relatively large trading volume is actually 1.1% less likely to place than a horse with almost identical pre-race odds that did not receive such volume. The

---

<sup>3</sup>For small fields there are only two places, mid-size fields have 3 places, and in large handicap races (>16) there are four places.

average trade is, however, more informative prior to races in the win market where more precise predictions are made. A horse that receives relatively large trading volume is 6.2% more likely to win than its low volume but same price equivalent.

The picture is more consistent when we consider trading during races. A horse that receives a disproportionate amount of bets is 9.3% more likely to win and 18.9% more likely to place than horses trading at similar prices but without the same trading volume. In other words, while the average trade is only slightly informative regarding fundamentals when information is stale, it is substantially more predictive of fundamentals when information is fresh. Putting the two time periods together, there is evidence to suggest that speculative trade is at least conducive and certainly not unfavourable towards market efficiency. Perhaps the results suggest that Glosten and Milgrom's (1985) model of trading—albeit with different proportions of informed traders prior to and during information events—dominates the models of Milgrom and Stokey (1982), De Long *et al.* (1990) and Avery and Zemsky (1998) in our market.

This brings us to the external validity of our results. On the debit side, the most attractive property of the betting market is also its greatest weakness. As each market is short-lived and fundamentals revealed before the day is out, positive-feedback traders of the type modelled in De Long *et al.* (1990) are unlikely to be found. These traders know that a stock is over-valued but having observed that sentiment has taken over a section of noise traders they too buy stock in the expectation of capitalising on short-run price movements and exiting prior to any ultimate correction. Due to the brevity of trading on the horse-racing we study, it is unlikely that traders buy (short) assets that they know are over (under) valued.

On the credit side, we have a pure speculative market without any of the other trading motivations such as hedging or risk-sharing found in financial markets. While betting is a leisure activity for most, there are a number of informed, experienced professionals present in this market. Moreover, the incentives for efficiently predicting winners are clear. More than 6 billion GBP is staked on the horse races that we study. Finally, and perhaps most importantly, the market structure mirrors the dominant market structure of financial markets (Parlour and Seppi, 2008). The limit order book operated by Betfair allows for the backing (buying) and laying (shorting) of horses, the submission and cancellation of limit and market orders, and the use of algorithmic trading strategies.

It is this market structure that distinguishes this paper from earlier work on the efficiency of

betting markets. Researchers have long noted that fundamentals are revealed in horse race betting. Griffith (1949), Ali (1977), Snyder (1978), Busche and Hall (1988), Thaler and Ziemba (1988) and Snowberg and Wolfers (2010), among many others, analyse betting returns in U.S. pari-mutuel markets. In these markets, the payout on each horse in the case of a win is inversely proportional to the amount wagered on that horse. In other words, prices cannot reflect information *without* trade. The existence of a positive favourite-longshot bias—where returns to betting on favourites exceed those of betting on longshots—suggests that trade goes some way but not quite far enough to incorporating all information into asset prices. On an alternative path, Shin (1993) and Vaughan Williams and Paton (1997), among others, have looked at U.K. bookmaker markets.<sup>4</sup> Here, prices can reflect fundamentals without any trade, as bookmakers can update their quotes irrespective of trading volumes, but the patterns of bets taken by bookmakers are not typically revealed. The betting exchange—by replicating the dominant financial market structure and providing the transparency necessary to construct detailed data on the volume of orders—allows for a clean test of the role of speculative trade in market efficiency.<sup>5</sup>

## 2 Data

The focus of our study is the Betfair betting exchange. The exchange is the largest of its kind in the world with the majority of its customers based in the U.K. Prior to the development of this exchange and others of its kind, the U.K. betting market was comprised of bookmakers and the Tote, a pari-mutuel pool. On the exchange, bettors can bet on or bet against a given horse, and can also submit both market orders and limit orders. Market orders execute with limit orders already in the book, and limit orders sit in the book until an offsetting market order arrives. Prior to the inception of the exchanges, it was not possible for bettors to take short positions or submit limit orders.

Prices on the exchange are quoted in the form of odds including the stake. For example, if a bettor backs (buys) a horse at 3, then 2 pounds will be returned for every pound staked if the horse wins. If a bettor lays (shorts) a horse at 4, he/she is liable to pay 3 pounds to his/her

---

<sup>4</sup>Detailed surveys of the literature, with a focus on pari-mutuel and bookmaker markets, can be found in Sauer (1998), Vaughan Williams (2005) and Hausch and Ziemba (2008).

<sup>5</sup>Smith *et al.* (2006) also study the favourite-longshot bias on Betfair, finding a slight positive bias, but do not consider the role of trade in market efficiency.

counterparty for every pound accepted if the horse wins. The pricing grid ranges from 1.01, which corresponds to 1 pence for each pound staked, to 1000, which corresponds to 999 pounds for each pound staked. There is no margin trading, at least not for the ordinary punter, as all liabilities must be lodged with the exchange prior to any order being submitted.

To illustrate the format of trading on the exchange, we present two screenshots of the Betfair limit order book in Figure 1. The screenshots are taken from the win market for the 14.00 race at Chepstow on the 30th September 2014. This race is outside of our data-set and is used for illustrative purposes. Market order traders can back a horse (take a long position) on the left-hand-side of the book, or lay a horse (take a short position) on the right-hand-side. The best three back (long) and lay (short) quotes are displayed with the volume available at each price indicated just below. In the top panel we display the screen prior to the race with the in-running screen in the bottom panel. As the race progresses, certain horses become more favoured and others drift out of contention. The favourite, Parish Business, narrowed from 3.8 prior to the race to 2.1 in our in-running example. On the other hand, Significant Move, priced at 30 beforehand, drifted out of contention as no limit order traders were offering a price for a market order trader to lay (short) this horse during the race.

**14:00 Chepstow** Live Radio Tote Multiples  
 Tue 30 Sep | 2m4f Mdn Hrd

Win Place AvB

Going In-Play Rules Betfair SP Timeform Matched: GBP 270,768 Refresh

14 selections Back all Lay all

5	<b>Parish Business</b> Aidan Coleman	3.7 £1621	3.75 £1512	3.8 £1012	3.85 £597	3.9 £2296	3.95 £1503
3	<b>Laser Hawk</b> Paul Moloney	4.2 £546	4.3 £711	4.4 £528	4.5 £1499	4.6 £563	4.7 £1026
6	<b>Primo Capitano</b> Ian Popham	4.5 £544	4.6 £419	4.7 £171	4.8 £418	4.9 £2227	5 £436
10	<b>Thespis Of Icaria</b> Nick Scholfield	6 £697	6.2 £576	6.4 £228	6.6 £210	6.8 £1378	7 £590
2	<b>Hold The Fort</b> Richard Johnson	19.5 £95	20 £368	21 £64	22 £107	23 £90	24 £87
8	<b>Significant Move</b> Conor O'Farrell	28 £42	29 £118	30 £22	32 £97	34 £59	36 £13
4	<b>L Frank Baum</b> Robert Dunne	40 £100	42 £11	44 £18	48 £18	50 £39	55 £56
12	<b>Dancing Olqa</b> Liam Heard	34 £13	36 £38	38 £29	40 £11	42 £64	55 £18
7	<b>Ruperra Tom</b> Richie McLernon	100 £13	120 £14	140 £10	150 £19	170 £17	200 £11
1	<b>Girevole</b> Tom O'Brien	170 £15	230 £11	270 £16	390 £12	400 £17	550 £10

(a) Pre-Race

**14:00 Chepstow** Live Radio Tote Multiples  
 Tue 30 Sep | 2m4f Mdn Hrd

Win Place AvB

In-Play Rules Betfair SP Timeform Matched: GBP 628,074 Refresh

14 selections Back all Lay all

5	<b>Parish Business</b> Aidan Coleman	2.02 £302	2.06 £49	2.1 £1089	2.22 £53	2.24 £41	2.26 £20
3	<b>Laser Hawk</b> Paul Moloney	3.8 £30	3.85 £81	3.9 £12	4.8 £13	5 £10	6.2 £58
6	<b>Primo Capitano</b> Ian Popham	27 £17	32 £94	34 £10	65 £11	220 £10	900 £9
10	<b>Thespis Of Icaria</b> Nick Scholfield	5.3 £25	5.5 £11	5.9 £16	9.2 £31	9.6 £29	11 £12
2	<b>Hold The Fort</b> Richard Johnson	7.8 £11	8 £16	8.6 £14	12.5 £10	13.5 £11	18 £12
8	<b>Significant Move</b> Conor O'Farrell	34 £13	190 £40	1000 £575			
4	<b>L Frank Baum</b> Robert Dunne	24 £10	34 £12	44 £11	320 £10	990 £5	
12	<b>Dancing Olqa</b> Liam Heard	27 £35	29 £14	60 £11	140 £10	200 £10	730 £10
7	<b>Ruperra Tom</b> Richie McLernon	12.5 £21	13 £25	20 £17	44 £39	90 £14	190 £10
1	<b>Girevole</b> Tom O'Brien	100 £27	170 £12	1000 £21			

(b) In-Running

Figure 1: Examples of the Betfair limit order book both prior to a race, in the top panel, and during a race, in the bottom panel. These screenshots are taken from the win market on the 14.00 race at Chepstow on the 30th September 2014. Punters can back (buy) horses by clicking on quotes on the left-hand-side of the book, and lay (short) horses by clicking on quotes on the right-hand-side. The best three quotes for back and lay bets with associated volumes underneath are displayed. Bettors can also access the book algorithmically through the Betfair Application Programming Interface.



We obtained data on 9,562 U.K. horse races from Fracsoft, a third-party provider of Betfair limit order book data. This is the full database of U.K. races between the 24th March 2013 and the 19th March 2014 inclusive. This data comprises second-by-second quotes of the best back (buy) and lay (sell) quotes, the last matched price, and the total matched volume on that bet at that second. This data includes both the win and place markets, and both pre-race betting and ‘in-running’ trading.<sup>6</sup> We concentrate on the last 20 minutes of trading prior to each race, calculating that 80.4% of pre-race trading, by volume, takes place in this window. Certain markets, trading periods, and horses are missing from the database, particularly for the place market as it goes in-running. In total, however, we have approximately 225 million observations of the Betfair limit order book spread over a year.

This data-set is then merged with data provided by Betfair on the winners and those that place in each race.<sup>7</sup> Approximately 10% of the Fracsoft data could not be matched with Betfair win data and was therefore dropped from the analysis.<sup>8</sup>

Our first step is to sum up all orders to give a picture of the volume on each horse. We separately consider each horse in each market (win or place), and in each time period (pre-race or in-running). In general, there is more trade in the win market and prior to racing. For example, the average horse is subject to 53,600 GBP of trading volume in the win market prior to races, but only 17,780 GBP in the same market during races. Of course the latter can be explained by the shorter duration of in-running betting.

We are primarily interested in whether or not a given horse received a disproportionate amount of volume. But what is a disproportionate amount? This has to be measured relative to other horses in the same race, as certain races may be subject to more bets than others. With this in mind, we calculated the median volume for bets within each race. This was done separately for win/place markets and pre-race/in-running trading periods. We then created an indicator variable for whether or not a given bet received more than the median volume within its race. This cut-off

---

<sup>6</sup>The market is suspended for one second just before the race begins, thereby cancelling all pre-race quotes: consequently quotes must be submitted afresh when the race begins.

<sup>7</sup>This data can be found at <http://www.betfairpromo.com/betfairsp/prices/index.php>.

<sup>8</sup>We analysed the Betfair data and found that there were no significant differences, in terms of race type and betting price, between the missing and non-missing data. We therefore have little reason to believe that there is a selection effect which may bias our upcoming results. Furthermore, we also observed that if one bet is missing from a race, all bets are missing from that same race. It is therefore not the case that we have a disproportionate number of favourites/longshots in our data-set.

varies substantially across races. The average cut-off is 30,848 GBP in the win market prior to races, and 8,272 GBP during races. The lowest median volume within a race is 0 GBP in pre-race betting and the highest is 610,227 GBP. In other words, it is possible to be classified as a horse that received a large amount of volume after only a few small bets, and it is also possible to be classified as a horse that received little volume even after many large bets. It depends on the volume received by the horse's competitors in that race, in that market, and in that trading period.<sup>9</sup>

It is important to note the differences in the types of horses/bets that receive more volume. In Table 1 we summarise the implied win probabilities for horses in our data. If the Betfair odds are  $X$  (including the stake), then the implied win probability is  $1/X$ . The implied win probability can be thought of as the price of an Arrow-Debreu security, paying 1 if the horse wins and 0 otherwise. Bets are broken down by market (win and place) and trading period (pre-race and in-running). Bets are also broken down into those that received greater than the median volume within their race and those that did not. We can observe that there is a clear pattern of higher volume for favourites. To give an example, the average implied win probability in the win market prior to races was 0.186 for high volume horses and 0.048 for low volume horses. Similar results emerge for the place market and in-running trading.

---

<sup>9</sup>We also designated the cut-off as the mean volume within each race. This tends to shift the cut-off slightly to the right, due to the positive skew of volume. However, the results were qualitatively similar to those that we present in the paper and therefore we did not include them here.

<b>Table 1: Implied Win Probability</b>					
<b>Panel A: Pre-Race</b>	Obs	Mean	Std. Dev.	Min	Max
Win Market					
$\leq$ Median Volume within Race	44,138	.048	.044	.001	.49
$>$ Median Volume within Race	41,532	.186	.126	.001	.975
Place Market					
$\leq$ Median Volume within Race	44,410	.181	.115	.001	.972
$>$ Median Volume within Race	39,733	.459	.169	.02	.982
<b>Panel B: In-Running</b>	Obs	Mean	Std. Dev.	Min	Max
Win Market					
$\leq$ Median Volume within Race	43,363	.055	.053	.001	.71
$>$ Median Volume within Race	40,979	.28	.2	.001	.99
Place Market					
$\leq$ Median Volume within Race	32,069	.314	.178	.002	.99
$>$ Median Volume within Race	39,198	.645	.203	.005	.99

**Summary statistics on horse implied win probabilities.** Implied win probability is defined as  $1/X$  where  $X$  is the weighted average odds. Implied win probabilities are equivalent to the price of Arrow-Debreu securities which pay 1 if the horse wins and 0 otherwise. Horses are divided into those that received greater than the median amount of volume in their race and those that did not. The statistics are also divided up by win/place market and pre-race/in-running trading periods. Across all of these sub-samples, more heavily favoured horses with higher implied win probabilities tend to attract greater betting volume.

This complicates any analysis because of the favourite-longshot bias observed by Smith *et al.* (2006) for this betting exchange. In this bias, favourites (longshots) win more (less) often than their odds suggest. Therefore, if we simply compared the returns of high and low volume bets, we could confound the predictive power of trading volume with the favourite-longshot bias.

To illustrate this point, consider the plots in Figure 2. We plot the actual win probabilities of horses/bets against the implied win probabilities described in Table 1. Implied win probabilities are rounded to two decimal places just for these graphs, in order that we have sufficiently large buckets to give smooth average win probabilities. Win and place bets are displayed separately

and pre-race (in-running) bets are displayed in the top (bottom) panel. Most bets in the win market can be found at low implied win probabilities, while there is a more even spread in the place market.

For pre-race trading, implied win probabilities closely correspond to actual win probabilities. A 45 degree line is added for comparison. When it comes to in-running trading, however, there is a pronounced favourite-longshot bias with favourites (longshots) winning much more (less) than their odds imply. This necessitates the comparison of high/low volume bets with the same or very similar implied probabilities, particularly when we look at in-running betting.

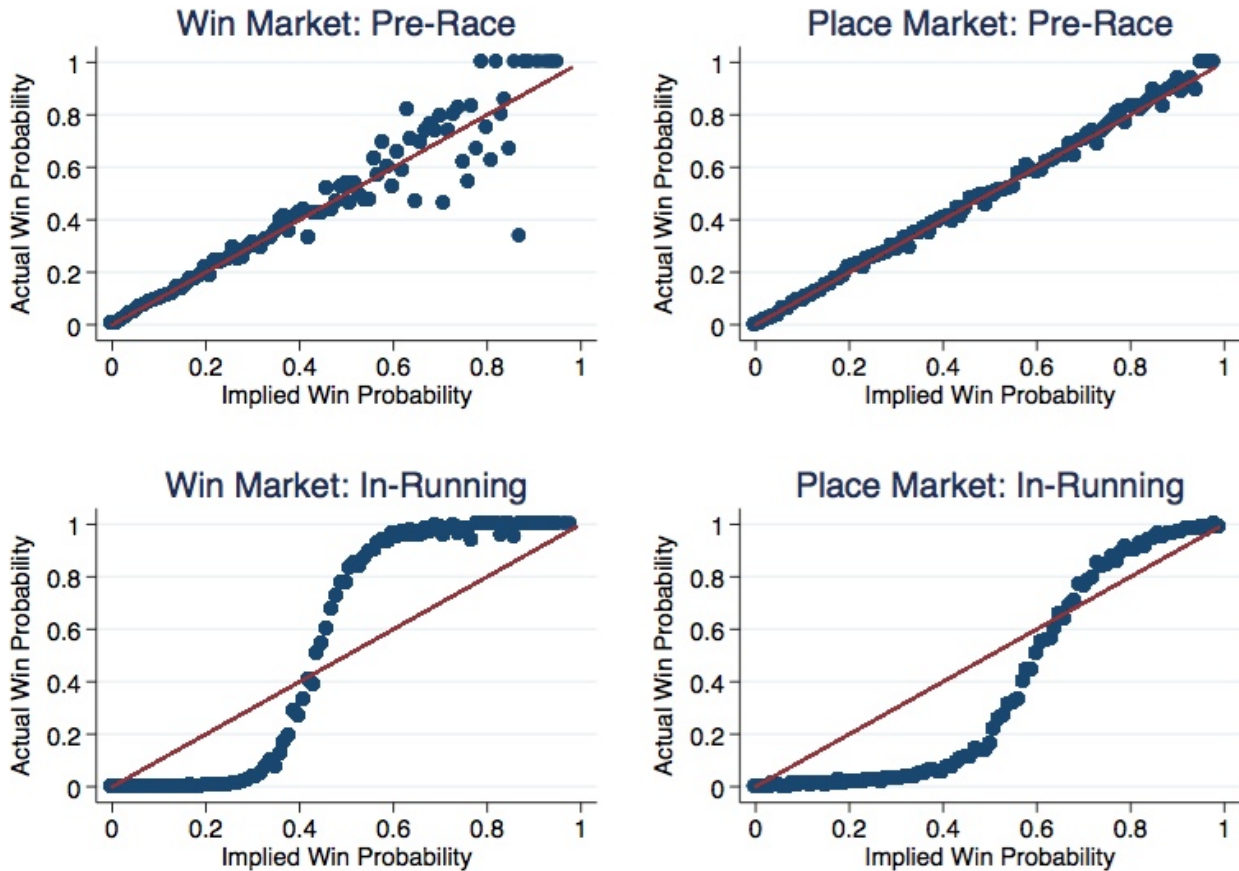


Figure 2: Plots of implied win probability—inferred from the betting odds for each horse—against actual win probability. Data for win and place markets are displayed separately. Pre-race betting is in the top panels and in-running betting is in the bottom panels. The implied win probability is rounded to 2 decimal places to create sufficiently large buckets for comparison with actual win probabilities. A pronounced favourite-longshot bias, where favourites (longshots) win more (less) often than their odds suggest, can be observed during in-running betting. A 45 degree line representing bias-free prices is added for comparison.

## 3 Analysis

### 3.1 Main Results

The aim of our analysis is to compare the win probabilities of horses that received greater volume with those that received less. If the win probabilities are higher for high volume bets after controlling for implied win probability (prices), this would suggest that speculative trade disproportionately targets mispriced quotes and valuations and is therefore conducive to market efficiency. In other words, we are suggesting that if speculative trade is conducive to market efficiency, a more *inefficient* contract should be more *heavily* traded. An alternative view is that if speculative trade is conducive to market efficiency, we might expect that more *heavily* traded contracts are more *efficient*. These ostensibly opposing views are explained in terms of the precise time at which the price or mispricing is measured.

If the price is measured *before* trade or *during* trade, as it is in our paper with the weighted average transaction price, we would expect higher levels of trade in mispriced contracts. This is because informed speculators identify the mispriced contracts and attempt to profit by trading more in these contracts. We will find evidence consistent with this behaviour.

If on the other hand the price is measured *after* trade, it is possible that the more heavily traded contract will have less mispricing and be more efficient. This is because speculative traders have impounded more information into the price of the heavily traded contract via their trade, and therefore increased its efficiency relative to the less heavily traded contract. It is important to note, however, that it is also possible that there is zero correlation between trading volume and price efficiency when prices *after* trade are used. This would occur if the lightly traded contract was already efficient and therefore, unlike the heavily traded contract, did not require trade to make it efficient. Along the same lines, even if we find that heavily traded contracts are less efficient than lightly traded contracts, based on prices *after* trade, this does not establish that speculative trade is detrimental to market efficiency. Informed traders may have eliminated some but not all of the mispricing. The important point when looking at price paths is whether the degree of mispricing declined, from the initial price to the last price, with the level of trading. We have a series of tests later in the paper which capture this point.

At this stage we can say that, as we are using prices *during* trade, a positive correlation between mispricing and trading volume, as we will find, would suggest that speculative trade is conducive

to market efficiency.

This is also a good time to make another clarification. In this paper we analyse whether high volume trading indicates that a horse is *more* likely to win than its price implies. In other words, we are only identifying the extent to which speculative trade can predict the winners of races, and not the extent to which it predicts the losers. This approach is chosen because it allows us to capture the informativeness of both market orders and limit orders. In an earlier version of the paper, we focused instead on the net volume on a horse: the volume of back orders by market order traders minus the volume of lay orders by market order traders. This earlier approach allowed us to capture the informativeness of short positions as well as long positions, but necessitated the neglect of limit orders. We find similar results with both approaches and therefore we only present the former in the paper.

We begin our analysis by estimating a linear probability model. We use this type of model, rather than a logit/probit model, in order to facilitate coefficient comparisons with our subsequent nearest neighbour matching approach. We regress an indicator variable equalling 1 if the bet won (the horse won or placed, depending on the market) on the implied win probability of the bet, and an indicator variable equalling 1 if the bet received greater than the median volume within its race. In Table 2 the results of this regression are displayed for win and place markets and pre-race and in-running periods separately. We cluster heteroskedasticity-consistent standard errors at the race level.<sup>10</sup>

---

<sup>10</sup>An alternative to the linear probability model with clustering at the race level is the conditional logit model. This model assigns observations to groups, in our case races, and the likelihood is calculated relative to the group. Our main results in the paper—that trade is in general more informative during races—are also backed by the results using the conditional logit model. However, we present only the results of the linear probability model in order to aid comparison with the nearest neighbour estimates to follow in this Section.

<b>Table 2: Regression Analysis</b>				
Dependent Variable: Win Indicator	1	2	3	4
Intercept	-0.001 (.001)	-0.004 (.001)	-0.088*** (.000)	-0.336*** (.001)
Implied Win Probability	0.997*** (.014)	1.027*** (.008)	1.61*** (.004)	1.279*** (.005)
> Median Volume within Race (Indicator)	0.0002 (.002)	-0.013** (.004)	-0.124*** (.001)	0.119*** (.003)
Market	Win	Place	Win	Place
Trading Period	Pre-Race	Pre-Race	In-Running	In-Running
$R^2$	0.129	0.187	0.687	0.578
No. of Clusters (Race-Market Combinations)	9,560	9,534	9,562	9,537
No. of Observations	83,942	84,110	83,170	71,235

**Regression analysis.** An indicator, equalling 1 if the bet won (the horse won or placed, depending on the market), was regressed on the implied win probability of the bet inferred from the weighted average odds, and an indicator variable equalling 1 if the horse/bet received greater than the median amount of volume within the race. The win (place) market is considered in regressions 1 and 3 (2 and 4) and pre-race (in-running) trading periods are considered in regressions 1 and 2 (3 and 4). Standard errors—heteroskedasticity-consistent and clustered at the race-market level—are in parentheses, and \*\*\* and \*\* indicates significance at the 0.1% and 1% level respectively.

The first point to note with the results in Table 2 is that the linear probability model fits the data much better prior to races, the differences in  $R^2$ s notwithstanding, as testified by the coefficient close to 1 for implied win probability. This is to be expected given the relationship between actual and implied win probabilities displayed in Figure 2. Prior to races in the win market, high volume bets are not significantly more likely to win, after controlling for implied win probabilities, than low volume bets. In place markets, high volume bets are actually 1.3% less likely to win than low volume bets. During races, high volume bets are less likely to win than their low volume equivalents in the win market but more likely in the place market. However, these results have to be interpreted with some caution, as the coefficients associated with the implied



win probability variable are comfortably above 1 for the in-running samples. This would generate some implausible predictions for win probabilities for extreme favourites, such as when the implied win probability is close to 1.

We therefore require a model that does not specify a functional relationship between implied and actual win probabilities. Using nearest neighbour techniques, we can match a high volume bet with implied win probability  $Z$  with a low volume bet with the same implied win probability. In other words, this rules out the favourite-longshot bias as a confound. Nearest neighbour matching is predominantly used in the policy evaluation literature as a technique to non-parametrically match treated individuals to non-treated control individuals. It is a form of selection on observables. Nearest neighbours are determined by measuring the Mahalanobis distance between individuals based on observable matching variables. In our case, we wish to match ‘treated’ bets, those with greater than the median volume in its race, with ‘control’ bets, those with volume less than or equal to the median in its race, and match them based on implied win probability. We can then determine whether the high volume bets win more often. If they do, this would suggest that trading volume is disproportionately concentrated in picking off mispriced quotes and valuations, which increases the efficiency of markets.

Our main reason for using the matching procedure, as alluded to earlier, is to avoid the well-known favourite-longshot bias. There is evidence in Vaughan Williams and Paton (1997) and Smith *et al.* (2006) that such a bias is found in bookmaker markets and on betting exchanges. There are a multitude of reasons for this—Ottaviani and Sørensen (2008) survey a number of these explanations—and to avoid confounding the effect of trading volume with these alternative explanations we use the matching procedure. In doing so, we can state that trading volume predicts fundamentals not simply because trading volume is concentrated in favoured horses. We are comparing the frequency with which a high volume bet with a given implied win probability wins, with the frequency with which a low volume bet with a very similar implied win probability wins. These high and low volume bets are therefore equally favoured, thereby ruling out the possibility that differences in win frequencies between the two are simply due to the favourite-longshot bias.

We begin by matching only on implied win probability. For this and all of the following matching procedures, we allow for treatment bets to be matched with more than one control bet and we match with replacement. In other words, a control bet can be matched with more than

one treatment bet and vice versa. As shown in Table 1, high volume bets have lower odds and higher implied win probabilities than low volume bets. The task therefore is to organise treatment and control groups in such a way that any differences in the matched pairs are insignificant. The nearest neighbour matching procedure does this. We conduct t-tests on the matched pairs of treatment and control bets, and find that, after matching, differences in implied win probability are insignificant even at the 10% level. This applies to both the win and place market and the pre-race and in-running trading periods. In other words, there is a large region of common support where both high and low volume bets with very similar or identical implied win probabilities can be found.

Once this matching has taken place, we can then compare the actual win probabilities of the matched bets. Analysis for the pre-race and in-running periods can be found in Table 3. Looking at the pre-race data in estimation 1 of Panel A first, we find that high volume bets are insignificantly more likely to win than their low volume equivalents in the win market. In the place market results in estimation 4 of Panel A, high volume bets are actually 1.4% less likely to win. The picture changes dramatically when we look at the in-running trading period in estimations 1 and 4 of Panel B. High volume bets are 9.3% more likely to win than their low volume equivalents in the win market with significance at the 0.1% level, and 13.3% more likely to win in the place market with significance also at the 0.1% level. Put simply, trade is highly predictive of fundamentals when information is arriving.

**Table 3: Nearest Neighbour Analysis**

**Panel A: Pre-Race**

Dependent Variable: Win Indicator	1	2	3	4	5	6
> Median Total Volume within Race (Indicator)	-0.002 (.011)	0.06. (.035)	0.062. (.035)	-0.014* (.006)	-0.003 (.022)	-0.011 (.022)
Market	Win	Win	Win	Place	Place	Place
Matching on Implied Win Probability	Yes	Yes	Yes	Yes	Yes	Yes
Matching on Median Volume within Race	No	Yes	Yes	No	Yes	Yes
Bias Adjustment for Continuous Matching Variables	No	No	Yes	No	No	Yes
No. of Observations	83,942	83,942	83,942	84,110	84,110	84,110

**Panel B: In-Running**

Dependent Variable: Win Indicator	1	2	3	4	5	6
> Median Total Volume within Race (Indicator)	0.093*** (.015)	0.113*** (.001)	0.114*** (.001)	0.133*** (.004)	0.19*** (.006)	0.189*** (.006)
Market	Win	Win	Win	Place	Place	Place
Matching on Implied Win Probability	Yes	Yes	Yes	Yes	Yes	Yes
Matching on Median Volume within Race	No	Yes	Yes	No	Yes	Yes
Bias Adjustment for Continuous Matching Variables	No	No	Yes	No	No	Yes
No. of Observations	83,170	83,170	83,170	71,235	71,235	71,235

Nearest neighbour analysis. Pre-race analysis is in Panel A with in-running analysis in Panel B. The average win indicator is compared for bets where the horse/bet received greater than the median amount of volume within the race, the ‘treated’, with those that did not, the ‘control’. Treatment and control bets are matched at various stages on implied win probability and the median volume within their race. In rows 3 and 6 of each panel we also adjust for biases that arise in estimates when there is more than 1 continuous matching variable (see Abadie and Imbens, 2011). Heteroskedasticity-consistent standard errors are in parentheses and \*\*\*, \* and . indicates significance at the 0.1%, 5% and 10% level respectively.

In the remainder of Table 3, we conduct a series of robustness exercises. First, we add an additional matching variable, the median total volume within the race. The reasoning for this is as follows. An ideal test would be to compare the win probabilities of one high and one low volume bet with identical implied win probabilities from exactly the same race. Unfortunately there are few bets with the same implied win probabilities within the same race. Our next best option is to find treatment and control bets from races with similar betting patterns. We can judge race similarity by the location of the cut-off, the median volume within the race. In other words, how much volume did it take to be classified as a high volume bet? We present the results of these exercises in columns 2 and 5 of both panels in Table 3.<sup>11</sup> Our final robustness exercise is to adjust for a bias in the matching estimators when there is more than one continuous matching variable. Following Abadie and Imbens (2011), the results of these estimations are in columns 3 and 6 of both panels in Table 3.

We find a similar pattern across all specifications. The average trade is slightly informative prior to races, and then only in the win market, but is very informative during races, particularly in the place market. Using our preferred matching procedure in columns 3 and 6 of panel A, we find that high volume horses are 6.2% more likely to win and 1.1% less likely to place than their low volume equivalents in betting prior to races. In betting during races in columns 3 and 6 of panel B, the roles reverse, with high volume horses 11.4% more likely to win and 18.9% more likely to place than their low volume equivalents. In other words, the average speculative trade has only a little to contribute when information is stale and markets are perhaps already efficient, but plays a key role in the incorporation of new information into asset prices.

To illustrate and summarise our results, we present Figure 3. This is a replication of Figure 2 but here bets are broken down into those that received high volume in that race, defined as above the median, and those that did not. As in Figure 2, pre-race bets are presented in the top panel and in-running bets in the bottom panel. Win and place market bets are displayed separately. We observe little difference in the win probabilities of high and low volume bets for each implied win probability prior to races, but there are pronounced differences during races. High volume bets

---

<sup>11</sup>As a further robustness exercise not displayed in the paper, we also matched treatment and control bets on the number of runners in the race, the distribution (specifically the mean and standard deviation) of other prices in the same race, and the total volume of bets in the race. Our results were qualitatively similar with these additional matching variables. Much of the information contained in the prices of other bets and the total value of the pool of bets is already in part captured by the price of the bet in question and the median volume in the race.

pay out much more frequently than their low volume equivalents during races. In other words, speculative trade is much more predictive of fundamentals as information arrives. This is most evident for implied win probabilities greater than 0.5, which more often than not are found in the place market.

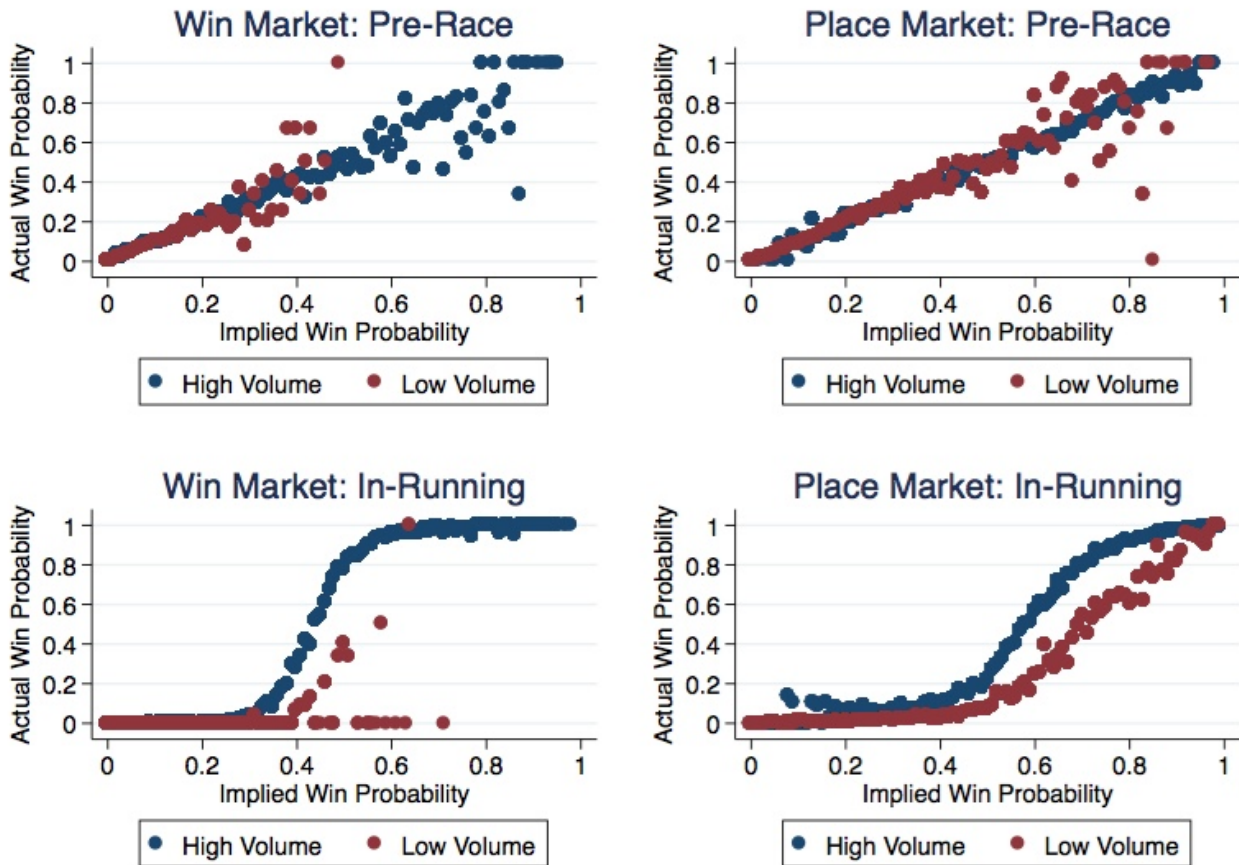


Figure 3: Plots of implied win probability—inferred from the betting odds for each horse—against actual win probability. For this figure we break down bets into those that received high volume, defined as above the median volume in that race, and those with low volume. The implied win probability is rounded to 2 decimal places to create sufficiently large buckets for comparison with actual win probabilities. Pre-race betting is in the top panels and in-running betting is in the bottom panels. While there is little difference between the win probabilities of high and low volume bets prior to races, high volume bets pay out much more frequently, and are therefore much more predictive of fundamentals, during races.

We should clarify that we are using an absolute measure of mispricing rather than a relative measure. Therefore no distinction is made in our analysis between a mispricing of 2% when the horse has a 50% chance of victory, and a similar mispricing of 2% when the price implies that the horse has only a 5% chance. This approach is particularly important when we consider differences across win and place markets. Recall that in almost all circumstances win market prices (implied win probabilities) will be lower than place market prices as fewer horses win than place. Therefore if win market and place market bets were equally informative about fundamentals in absolute terms, win market bets would be more informative in relative terms. This observation makes the greater informativeness of win market trade prior to races in absolute terms all the more striking, and puts the inferior informativeness of win market trade during races, again in absolute terms, into perspective.

### 3.2 Additional Exercises

One issue is that our analysis does not show that speculative trade makes subsequent prices more efficient. We have shown that traders disproportionately trade in mispriced assets but it does not necessarily follow that subsequent prices are more efficient. A mispriced quote or valuation picked off by a speculative trader may be replaced by a similarly mispriced quote or valuation in the next second.

One way to examine whether speculative trade makes subsequent prices more efficient is to look at the first quoted price before trade took place and the last quoted price after trade took place, to see whether the degree of mispricing has declined between the two stages. To put more flesh on our approach, suppose there are three horses A, B, and C. A and C have a similar fundamental probability of winning while B has a lower fundamental probability of winning. Further, suppose that prior to trading, B and C are accurately priced but A is priced similarly to B when its fundamental value is actually closer to C. An informed speculator will buy bets on A, driving the price of A toward C and away from B. In so doing, the speculator causes A to receive an abnormal amount of trading volume. If we match treatment and control bets *before* trade, A, the high volume bet, will be matched with B, a low volume bet, as their prices are similar, and yet A will win more often than B, thereby indicating that there was a mispricing. If we match treatment and control bets *after* trade, A, the high volume bet, will be matched with C, a low volume bet, as their prices are now similar and both bets will win approximately as frequently, thereby indicating

that there is now no mispricing. In other words, the degree of mispricing between treated bets, those that receive abnormal trading volume, and control bets, those that do not, should be greater *before* trade than *after* trade if speculative trade is indeed conducive to market efficiency.

With this in mind, we conduct the analysis displayed in Table 4. We will begin by discussing the pre-race results in Panel A. We designate a bet as treated if it received more than the median volume in its race, and designate a bet as a control bet if it did not. In three separate estimations, we match these treatment and control bets on three different implied win probabilities: the first quoted price 20 minutes before the race (*before* trade), the trading volume weighted average transaction price (*during* trade), and the last quoted price just as the race is about to begin (*after* trade). The results in relation to *during* trade have already been displayed in Table 3. We then compare the win indicator for these treatment and control bets. We match treatment and control bets only on the basis of implied win probability for space reasons, but our results are qualitatively similar when we include additional matching variables, such as the median volume within the race, as we did in Table 3.

For the pre-race period, we find evidence to suggest that speculative trade does lead to more efficient subsequent prices in the win market. If we match on the first quote (*before* trade), treated bets, those that subsequently receive more trading volume, win 1.5% more often than control bets with similar prices which subsequently receive lesser volume. When we match the bets on the basis of the trading volume weighted average price (*during* trade), or the last quoted price (*after* trade), treated bets win insignificantly more/less than control bets. In other words, it would appear that prices begin the trading window somewhat inefficient, and then speculative trade drives the prices toward their fundamental value so that, by the time trading has ceased, this particular inefficiency has disappeared. However, the results are more mixed for the place market. Mispricing declines and then overshoots somewhat from *before* trade to *during* trade (1.2% to -1.4%). Then, mispricing is actually higher *after* trading than it was *before* trading (4.9% versus 1.2%). This supports the view that place market trade prior to races is not particularly conducive to market efficiency.

In Panel B of Table 4 we repeat the analysis for the in-running period. The first quoted price for this analysis is calculated as the race begins, and the trading volume weighted average price corresponds to our earlier estimations in Table 3. For the in-running period, there is no analysis of quotes *after* trade, as by this time the race has been completed and fundamentals have been revealed. In other words, after in-running trade, there is by definition no mispricing. We



find that the degree of mispricing is much greater *before* trade than *during* trade. Taking the win market analysis in estimations 1 and 2 of panel B as an example, treated bets—which subsequently receive abnormal trading volume during the race—win 19.7% more frequently than control bets with similar pre-trade prices which do not subsequently receive such large volume. When we match on the basis of prices *during* trade, however, this difference is down to 9.3%. Similar results are found when we look at the place market in columns 4 and 5 of Panel B, as mispricing declines from 49.9% to 13.3%. This suggests that as information arrives, informed speculators profitably drive this information into prices.

**Table 4: Mispricing Before/During/After Trade**

<b>Panel A: Pre-Race</b>						
Dependent Variable: Win Indicator	1	2	3	4	5	6
> Median Total Volume within Race (Indicator)	0.015*	-0.002	-0.0001	0.012*	-0.014*	.049***
	(.006)	(.011)	(.009)	(.005)	(.006)	(.005)
Market	Win	Win	Win	Place	Place	Place
Before/During/After Trade	Before	During	After	Before	During	After
Matching on Implied Win Probability	Yes	Yes	Yes	Yes	Yes	Yes
No. of Observations	83,942	83,942	83,942	84,110	84,110	84,110
<b>Panel B: In-Running</b>						
Dependent Variable: Win Indicator	1	2	3	4	5	6
> Median Total Volume within Race (Indicator)	0.197***	0.093***	NA	0.499***	0.133***	NA
	(.002)	(.015)	NA	(.003)	(.004)	NA
Market	Win	Win	Win	Place	Place	Place
Before/During/After Trade	Before	During	After	Before	During	After
Matching on Implied Win Probability	Yes	Yes	NA	Yes	Yes	NA
No. of Observations	83,170	83,170	NA	71,235	71,235	NA

Analysis to establish whether market inefficiencies decrease after speculative trade. Pre-race analysis is in Panel A with in-running analysis in Panel B. The average win indicator is compared for bets where the horse/bet received greater than the median volume within the race, the ‘treated’, with those that did not, the ‘control’. Treatment and control bets are matched on one of three different implied win probabilities: either the first quoted price (*before* trade), or the trading volume weighted average price (*during* trade), or the last quoted price (*after* trade). There is no after trade price for the in-running period. Heteroskedasticity-consistent standard errors are in parentheses and \*\*\* and \* indicates significance at the 0.1% and 5% level respectively.

This brings us to our final analysis. Throughout the paper we have labelled a bet as a high volume bet if its total volume exceeded the median level within that race. But it is quite clear that traders will allocate their attention and resources unevenly across races. It is natural to think that an informed speculator will trade more in races where they expect that mispricing and therefore potential profits are greater. If we want to know the effect of speculative trade on mispricing, we should allow for the fact that speculators are not simply selecting which horse to back over other horses in a given race but are also choosing in which race to bet.

With this in mind, in our final analysis we utilise an alternative classification of our treatment and control bets. We designate a bet as ‘treated’ if it received greater trading volume than the median bet across *all races* and designate the bet as a ‘control’ bet otherwise. Win and place market bets are considered separately: thus a treated bet in the win market needs to have greater total volume than the median win market bet, and a treated bet in the place market needs to have greater total volume than the median place market bet. Treatment and control bets are once again matched on the weighted average transaction price (the implied win probability) to rule out the favourite-longshot bias as a potential confound. In contrast to Table 3, we do not include other race characteristics as matching variables. This is not only for space reasons but also because we want to allow as much flexibility as possible in the scope for informed traders to select across races. If we use race characteristics as additional matching variables, we are requiring that treatment and control bets come from very similar races which, in limiting to extent to which we allow informed trade to concentrate in certain races, defeats the purpose of this exercise.

Using our alternative classification, we find similar results to those already discussed in the paper. Take Panel A of Table 5, for example. Treated win market bets—those that received greater than the median volume of all horses in the win market—win 6.7% more frequently than control bets with similar prices but without the same level of volume. There is more fundamental information in trading volume in the win market than in the place market, at least in the window just before the race.

In Panel B of Table 5, we repeat our analysis for the in-running period. Our new in-running results also tally with the results of our earlier analyses. Speculative trade is predictive of fundamentals as races are run, particularly in the place market. For example, treated place market bets—those with greater than the median volume across all place market bets—win 19.9% more frequently than control bets. These control bets had similar weighted average prices but did not

receive such large volume during the race. To sum up, even when we acknowledge that informed speculators select across races and not just within races, our results look similar. Speculative trade is predictive of fundamentals, in particular when information is arriving and needs to be impounded into prices.

<b>Table 5: NN Analysis of Volume Across All Races (Pre-Race)</b>		
<b>Panel A: Pre-Race</b>		
Dependent Variable: Win Indicator	1	2
> Median Volume Across All Races (Indicator)	0.067. (.038)	0.012 (.021)
Market	Win	Place
Matching on Implied Win Probability (from Odds)	Yes	Yes
No. of Observations	83,942	84,110
<b>Panel B: In-Running</b>		
Dependent Variable: Win Indicator	1	2
> Median Total Volume Across All Races (Indicator)	0.114*** (.001)	0.199*** (.008)
Market	Win	Place
Matching on Implied Win Probability (from Odds)	Yes	Yes
No. of Observations	83,170	71,235

Nearest neighbour analysis, this time assigning ‘treatment’ and ‘control’ bets on the basis of volume across *all races*. Pre-race analysis is in Panel A with in-running analysis in Panel B. The average win indicator is compared for bets where the horse/bet received greater than the median amount of volume of bets across all races, the ‘treated’, with those that did not, the ‘control’. Treatment and control bets are matched on implied win probability. Heteroskedasticity-consistent standard errors are in parentheses and \*\*\* and . indicates significance at the 0.1% and 10% level respectively.

In the next section we will discuss our results more broadly in relation to the theoretical literature and also draw comparisons with empirical work in financial markets.

## 4 Discussion

Broadly, we find that trade is inconsistently predictive of fundamentals when information is stale but closely foreshadows terminal payoffs when information is arriving. Therefore taking the day's trading as a whole, speculative trade is on balance concentrated in mispriced assets and therefore, at least in part, drives asset prices towards their fundamental values. This would seem to support the Glosten and Milgrom (1985) picture of trade, where the trading population is a mixture of uninformed traders who trade randomly, thereby creating noise, and informed insiders who trade in the direction of fundamentals. The difference here is that the proportion of informed traders differs substantially across trading periods.

We should be clear that this does not rule out the presence of other types of traders. It is quite possible that there are bettors that herd, and bettors that hold erroneous beliefs about the prospect of a horse in a given race and who thereby buy (short) over (under) valued bets. For example, suppose there is an undervalued bet prior to a race. Perhaps informed bettors purchase a large quantity of this bet, knowing that the horse is more likely to win than quoted prices suggest. Perhaps at the same time there is another group of traders making the opposite bet and betting against this horse, believing the horse is over-rated. The informational content of the former's trades is therefore diluted by the erroneous beliefs reflected in the latter's trades. No matter. In essence, among all of the types of trader and trade that exist in this market, we are simply trying to get a picture of the fundamental information content of the average trade. This average trade is of slight information prior to races—perhaps unsurprisingly, as traders have had plenty of time to digest the public information available to them at that point—but is highly informative during races.

Though our approach is different, our interests in this paper are similar to those of the French and Roll (1986) paper. French and Roll wanted to find out whether elevated stock price volatility during trading hours was due to 1) the arrival of public information and therefore unrelated to the trading process, or 2) informed traders revealing their private information through trade, or 3) noise traders distorting prices from fundamentals. Exploiting the closing of New York exchanges due to a backlog of paperwork, they found that much of the volatility in stock prices could be attributed to the trading process, rather than to the scheduling of public news during exchange hours. Furthermore, in subsequent tests they attributed the majority of this volatility to informed traders revealing their information through trade, rather than noise traders distorting asset prices.

This matches with our finding that speculative trade is predictive of fundamentals, which implies that traders reveal their information and generate volatility through trade. Having said that, in our setting, much of the private information revealed through trade is information that would be revealed publicly seconds later.

This brings us to a discussion of more recent empirical results in financial markets. Kim and Stoll (2014) wanted to establish whether trading imbalances are predictive of the informational content of earnings announcements. Notwithstanding that earnings announcements are a somewhat short-term and noisy measure of fundamentals, they found that there was almost no relation between trading imbalances and the content, or surprise, in announcements. This seems to tally with our pre-race results. When information is stale, we cannot expect to garner much information on fundamentals by looking at volume of orders. On the other hand, there is a great deal of information contained in orders that take place during races. Moreover, trade seems to be the process by which some new information is incorporated into prices. This corresponds with work by Love and Payne (2008), who examine movements in foreign exchange rates around macroeconomic announcements. They estimate that up to a third of the informational content of these announcements is driven into prices via trade, rather than liquidity providers simply updating their quotes.

The second source of variation in our data is provided by the separation of win and place markets. More precise predictions are required in the former, much as an option market allows for the precise prediction of future stock price movements. For example, if an investor knows that a stock is significantly undervalued, it would make sense to leverage this information using an out-of-the-money call option, rather than simply buy the stock. While it has been noted since Black (1975) that informed investors may favour the option market due to this built-in leverage, recent work has examined relative volumes in option and equity markets to test the proposition. Roll *et al.* (2010) examine the option/stock volume ratio (the O/S ratio) around earnings announcements. They find that this ratio is higher prior to announcements, suggesting relatively more informed option market activity, and link this high ratio to post-announcement returns. Similar work has found that the O/S ratio is predictive of returns prior to mergers and acquisitions announcements (Augustin *et al.*, 2014) and in advance of bankruptcy filings (Ge *et al.*, 2014).

This corresponds with our results from the win and place markets. Prior to races, when information is stale, speculative trade is more predictive of fundamentals in the win market. During

races, when information arrives rapidly, speculative trade is informative in both markets but particularly in the place market. In other words, it would appear that traders need to leverage inside information and make more precise predictions when all public information is stale, but trade in a more coarse fashion when information is new and yet to be fully digested.

## 5 Conclusion

Trading volumes are at historical peaks in financial markets. While much of this trade may be for hedging and risk-sharing purposes, much is undoubtedly speculative in nature. The level of financial market activity has brought renewed calls for a financial transactions tax to curtail speculative activity. The main defence against such an imposition is two fold: transaction taxes will 1) reduce market liquidity, and 2) hinder market efficiency.

The problem with the second argument is that we have no way of assessing the true efficiency of most financial markets as fundamentals are not observed. In this paper, we therefore use horse race betting on an exchange—which shares many of the same features as financial limit order books—as a laboratory to assess the role of speculative trade in market efficiency. Crucially, the fundamental values of horse race bets are unambiguously revealed as each race ends.

We divide bets into those that received greater trading volume in that race, and those that did not. We can then compare the actual win probabilities, or fundamentals, for high and low volume bets with exactly the same implied win probabilities, inferred from the weighted average transaction prices. If the high volume bets win more frequently, trade is on average conducive to market efficiency.

We find that the predictive capacity of speculative trade varies substantially with the trading period. Prior to races, when information is stale and markets almost already efficient, the average speculative trade contains little information. The only significant trade-based information is in the win market where, as in option markets, precise predictions of fundamentals may allow for leveraging of small nuggets of information. When the race begins, however, speculative trade is an excellent predictor of fundamentals, particularly in the place market where more coarse predictions will suffice.

In short, speculative trade plays an important role in the incorporation of new information into asset prices. That said, the type of fundamental information that this trade reveals would be in

the hands of all traders, and not require much in the way of processing, just a few seconds later.

## References

- Abadie, A., Imbens, G., W., (2011) Bias-Corrected Matching Estimators for Average Treatment Effects, *Journal of Business and Economic Statistics* **29**, 1-11.
- Ali, M., M., (1977) Probability and Utility Estimates for Racetrack Bettors, *Journal of Political Economy* **85**, 803-815.
- Angrist, J., D., Pischke, J., S., (2009) *Mostly Harmless Econometrics: An Empiricist's Companion*, Princeton University Press.
- Augustin, P., Brenner, M., Subrahmanyam, M., G., (2014) Informed Options Trading Prior to M&A Announcements: Insider Trading? Working paper, <http://ssrn.com/abstract=2441606>.
- Avery, C., Zemsky, P., (1998) Multidimensional Uncertainty and Herd Behavior in Financial Markets, *American Economic Review* **88**, 724-748.
- Black, F., (1975) Fact and Fantasy in the Use of Options, *Financial Analysts Journal* **31**, 36-41, 61-72.
- Busche, K., Hall, C., D., (1988) An Exception to the Risk Preference Anomaly, *Journal of Business* **61**, 337-346.
- Chordia, T., Roll, R., Subrahmanyam, A., (2011) Recent Trends in Trading Activity and Market Quality, *Journal of Financial Economics* **101**, 243-263.
- De Long, J., B., Shleifer, A., Summers, L., A., Waldmann, R., J., (1990) Noise Trader Risk in Financial Markets, *Journal of Political Economy* **98**, 703-738.
- French, K., R., Roll, R., (1986) Stock Return Variances: The Arrival of Information and the Reaction of Traders, *Journal of Financial Economics* **17**, 5-26.
- Ge, L., Humphery-Jenner, M., Lin, T., C., (2014) Informed Options Trading Prior to Bankruptcy Filings. Working paper, <http://ssrn.com/abstract=2302888>.



- 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.
- Griffith, R., M., (1949) Odds Adjustments by American Horse-Race Bettors, *The American Journal of Psychology* **62**, 290-294.
- Hausch, D., B., Ziemba, W., T., (2008) *Handbook of Sports and Lottery Markets*, North Holland,
- Kim, S., K., Stoll, H., R., (2014) Are Trading Imbalances Indicative of Private Information? *Journal of Financial Markets* **20**, 151-174.
- Love, R., Payne, R., (2008) Macroeconomic News, Order Flows, and Exchange Rates, *Journal of Financial and Quantitative Analysis* **43**, 467-488.
- Milgrom, P., Stokey, N., (1982) Information, Trade and Common Knowledge, *Journal of Economic Theory* **26**, 17-27.
- Ottaviani, M., Sørensen, P., N., (2008) The Favorite-Longshot Bias: An Overview of the Main Explanations, *Handbook of Sports and Lottery Markets*, North Holland, edited by Hausch, D., B., Ziemba, W., T., 83-101.
- Parlour, C., A., Seppi, D., J., (2008) Limit Order Markets: A Survey, *Handbook of Financial Intermediation and Banking*, edited by Boot, A., W., A., Thakor, A., V., Elsevier.
- Roll, R., Schwartz, E., Subrahmanyam, A., (2010) O/S: The Relative Trading in Options and Stock, *Journal of Financial Economics* **96**, 1-17.
- Sauer, R., D., (1998) The Economics of Wagering Markets, *Journal of Economic Literature* **36**, 2021-2064.
- 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.

- Snowberg, E., Wolfers, J., (2010) Explaining the Favorite-Long Shot Bias: Is it Risk Love or Misperceptions? *Journal of Political Economy* **118**, 723-746.
- Snyder, W., W., (1978) Horse Racing: Testing the Efficient Markets Model, *Journal of Finance* **33**, 1109-1118.
- Thaler, R., H., Ziemba, W., T., (1988) Anomalies: Parimutuel Betting Markets: Racetracks and Lotteries, *Journal of Economic Perspectives* **2**, 161-174.
- Vaughan Williams, L., (2005) *Information Efficiency in Financial and Betting Markets*, Cambridge University Press.
- Vaughan Williams, L., Paton, D., (1997) Why is there a Favourite-Longshot Bias in British Racetrack Betting Markets? *The Economic Journal* **107**, 150-158.