The Distribution of Information in Speculative Markets: A Natural Experiment[*]

Alasdair Brown[†]

University of East Anglia

June 9, 2014

[†]School of Economics, University of East Anglia, Norwich, NR4 7TJ. Email: alasdair.brown@uea.ac.uk

Abstract

We use a unique natural experiment to shed light on the distribution of information in speculative markets. In June 2011, Betfair - a U.K. betting exchange - levied a tax of up to 60% on all future profits accrued by the top 0.1% of profitable traders. Such a move appears to have driven at least some of these traders off the exchange, taking their information with them. We investigate the effect of the new tax on the forecasting capacity of the exchange (our measure of the market's incorporation of information into the price). We find that there was scant decline in the forecasting capacity of the exchange - relative to a control market - suggesting that the bulk of information had hitherto been held by the majority of traders, rather than the select few affected by the rule change. This result is robust to the choice of forecasting measure, the choice of forecasting interval, and the choice of race type. This provides evidence that more than a few traders are typically involved in the price discovery process in speculative markets.

1   Introduction

There is a common perception that the benefits of trading in speculative assets are skewed towards those on the inside of corporations, or those that move in the same circles as the corporate elite. Corporate insiders trade with knowledge of the latest deals and projects (see the long history of insider-trading cases), while retail investors trade on the basis of stale information in the financial press (Tetlock (2011)), or second-rate tips on stock-picking shows (Engelberg *et al.* (2012)). Predominantly for reasons of analytical tractability, many of the canonical models in financial economics do little to disabuse us of this notion. For example, in Glosten and Milgrom (1985) the trading population is divided into informed traders, who receive a signal on the fundamental value of the asset, and liquidity/uninformed traders, who trade randomly.

Yet there is increasing evidence that 'ordinary' retail investors possess relevant information of their own. Kelley and Tetlock (2013) find that the orders of retail investors contain novel information on firm fundamentals. Kaniel *et al.* (2012) find that the trades of retail investors predict abnormal returns around earnings announcements. There is good reason to expect such

results; retail investors may be located in geographical proximity to the firm, and even find their employment in a publicly-listed firm. These two recent results lead us to suspect that information in speculative markets may be more evenly spread than previously thought.

To investigate this further, we analyse a natural experiment that took place on Betfair, a U.K. betting exchange, in 2011. Betfair operates a public limit order much like the standard financial exchanges, except that the assets traded are not stocks, bonds or derivatives, but bets on sporting events. A breakdown of the betting population on Betfair would reveal a miniature version of a standard financial market: the exchange houses algorithmic traders, brokers who work on behalf of clients, organisations that send analysts out into the field to gather information, firms using mathematical models to model sporting event outcomes and, of course, your ordinary bettor.

In June 2011, Betfair announced that, with immediate effect, traders who had racked up profits of more than GBP 250,000 over the lifetime of their account, estimated at approximately 0.1% of traders, would see all future profits (after commission) taxed at up to 60%. Betfair were not simply targeting those who had got lucky. Betfair excluded any single market which accounted for more than 50% of a bettor's profits and, to be eligible for this new tax, bettors must have been active in at least 1000 markets and therefore proved their informational advantage over time. For these most informed bettors, Betfair ceased to be an attractive venue overnight.

We investigate the effect of this unique intervention on the forecasting capacity of this market (our measure of the market's incorporation of information into the price), by focusing on 11,307 horse races in the U.K. and Ireland that straddled the change in terms on Betfair. We utilise bookmaker prices for the same races as our control market, which allows us to capture whether races naturally got easier or harder to forecast after the intervention. Races may vary in their predictability dependent on the season (jump or flat racing), the number of horses in the field, the grade (or class) of the race, and also the weather. This is why our control group - for which we are forecasting exactly the same races - is crucial.

If the forecasting capacity of the betting exchange declined significantly, even relative to our control market, we can infer that much of the information in this simple market had hitherto lain with a select few informed traders. On the other hand, if the forecasting capacity of the exchange shows little decline, we can infer that much of the information in this market lay with the majority of traders who were unaffected by the change in terms.

After the intervention, the forecasting capacity of the exchange actually fell by 0.7%, compared

to a decline of 0.5% in our bookmaker control market. This difference turns out to be a statistically insignificant. To check the robustness of our results, we used different measures of forecasting capacity, sampled forecasting capacity at different intervals, and checked a number of sub-samples. All of these procedures led us to the same conclusion. Based on our results, it would appear that the majority of information in this market had hitherto been held by the majority of traders, rather than the select few affected by the rule change. To put it another way, the evidence suggests that more than a few traders are typically involved in the price discovery process in speculative markets.

In addition to being the venue for this unique experiment, betting markets have long proved popular for assessing the informational efficiency of markets (see Vaughan Williams (2005) for a survey). This is predominantly due to two attractive properties of betting assets: they are short lived (unlike, for example, most stocks), and the final payoffs can be easily defined. Much of the literature's focus has been on the favourite-longshot bias, where returns from betting on favourites have regularly exceeded those of betting on longshots (see Ottaviani and Sørensen (2008) for a technical survey of explanations for the bias). Among other things, the bias has allowed researchers to quantify the proportion of 'insiders' within a market. Our distinct contribution is to establish the role that these insiders actually play in the informational efficiency of the market. In other words we are asking the question, 'how would a market perform without its best performers?' It is, to our knowledge, unheard of for a financial market to penalise informed trading regardless of whether the information is legally or illegally obtained. Betfair's intervention therefore provides a unique opportunity to examine the extent to which the success of markets - as capital allocation mechanisms and event predictors - is due to its most adept participants.

How do our results fit in with the existing literature on the distribution of trading skills within speculative markets? Coval *et al.* (2005) analyse the performance of a sample of individual traders, provided by a U.S. brokerage firm. They find that trading performance in previous years is a key predictor of subsequent trading performance, suggesting that trading skill (or lack of it) is persistent. Barber *et al.* (2014) analyse a large sample of Taiwanese day traders - defined as those that buy and sell the same stock within one day - over a 15 year period. They too find that trading skill is persistent, with the top 500 ranked traders in year t outperforming the loss-making thousands by 72.8 basis points per day in year $t+1$. Only the extreme right-tail of traders - with the most persistent ability to time the intra-day market - are profitable after trading costs are

included. This fits in with earlier work which ascribes much of the underperformance of individual investors to over-trading (Barber and Odean (2000)).

In common with the Barber *et al.* (2014) setting, we have an extreme right-tail of profitable traders: the approximately 0.1% caught up in the new tax. We do not have data on individual trader accounts, but instead isolate the impact of these traders by utilising a rule change which specifically targeted them. After the rule change, and after a subset of these traders appeared to vacate the market, the forecasting capacity of market prices saw only a negligible dip (relative to a control market). This suggests that profitability in speculative markets is perhaps more about an ability to time the market - and make accurate short-term price forecasts - than it is about accurately predicting future events and fundamentals.

The rest of the paper is structured as follows. In Section 2 we describe the exchange and the 'premium charge' that saw top bettors' profits taxed at up to 60%. In Section 3 we present the natural experiment and Section 4 concludes.

2   The Betting Exchange

Betfair have been running a public limit order book for bets on sporting, cultural and political events since June 2000. The majority of traders are based in the U.K.. As with most public limit order books, traders can submit market orders, which match up with offers already in the book, or submit limit orders, which sit in the book until an offsetting order arrives. The contribution of the exchanges is to allow bettors to bet on or against a particular event (e.g. a horse to win), in comparison to bookmaker markets where only betting on an event is permitted. Betfair is the leading betting exchange in the U.K., with its nearest competitor being Betdaq.

The exchanges derive their income from matching bettors who wish to take opposite positions on a particular event, and charging a commission on the winner of the bet. The commission on Betfair varies between 2% and 5%, with the lowest rates being charged on those bettors who have accumulated loyalty points through regular and voluminous betting. Betdaq also charges commission of between 2% and 5%.

On 8th September 2008, Betfair announced that it would be augmenting its commission system with a 'premium charge', payable by all bettors who had accumulated a net profit on their account and had bet on more than 250 events. This first premium charge stipulated that profitable bettors

must pay at least 20% of the profits in commission, or be faced with a charge that brought their charges up to 20% of their profits.[1] A punter, quoted in *The Guardian* (9/9/08), mocked Betfair's previous ad campaign run on the slogan 'Come to Betfair, where winners are welcome' by describing their new charge as 'Come to Betfair, where winners get screwed'.

The first premium charge was a precursor to a second higher premium charge introduced on 20th June 2011. Bettors who had generated more than GBP 250,000 in profits on their account, yet paid less than 40% of this in commissions - and bet in more than 1000 markets - would be subject to a higher tax, payable after commission at the end of each week's betting. To ensure they didn't punish the lucky bettor, any market which had generated more than 50% of a bettor's profits was excluded from the total profits calculation. Bettors who satisfied the above conditions would be taxed 40% on their future profits if they had paid commission of more than 10% of their profits, be taxed 50% on their future profits if they had paid commission of between 5% and 10% of their profits, or be taxed 60% on their future profits if they had paid commission of less than 5% of their profits. Betfair estimated that 0.5% of bettors were caught in the first premium charge, and that approximately 0.1% would be caught by the second charge. We have data on the period that surrounds the second premium charge.

The higher premium charge was greeted in the same manner as the first, with some punters advocating a shift toward Betdaq, Betfair's main rival (*The Guardian* (29/6/11)). To illustrate the difference the new premium charge could potentially make to successful bettor's profits, consider the following choices of odds on the winner of the 2.20pm race at Southwell on March 1st 2012. A bet on the favourite, Samsana, at 2.17pm, would have given odds of 1.17 on Betfair, 1.1 on Betdaq, and 1 on William Hill, a prominent bookmaker. In the case of a win, this would give profits of 1 on William Hill, 1.07 on Betdaq (after 2% commission), and potentially only 0.45 on Betfair (after 2% commission and the 60% premium charge is levied). Prior to the second premium charge, a Betfair bettor would have collected at least 0.93 in the case of a Samsana win (after the maximum 20% commission was charged). The example given is hypothetical and extreme, as a bettor could offset some of the charge with losses incurred on other races, but, nonetheless, gives some idea of the stark worsening in betting conditions for the most informed bettors on Betfair.

The premium charge was applied to all betting on the exchange. There would undoubtedly

---

[1]To clarify, if a regular bettor never lost they would likely pay 2% of their profits in commission (losses would bring this total closer to 20% as no commission is charged on losses).

have been a proportion of affected bettors with specialisms other than horse-racing. However, our results in the next section suggest there was, as we expected, an effect on horse-racing insiders. The 'sport of kings' is more closely linked to betting than any other sport, and indeed is partially funded by a levy on betting proceeds. Punters can use many techniques to gain an informational advantage. They can study the form, observe the horses in the parade ring, or speak to the 'connections' (trainers and owners) to get an inside line on a horse's potential. The proportion of insiders in horse-racing has been estimated by Shin (1993) to total 2% of the betting population (from a sample of 136 races in July 1991), and similarly estimated at 1.9% by Vaughan Williams and Paton (1997) using a larger data-set of 510 races from 1992. The 0.1% affected in our study, bearing in mind that some focus on other sports, could therefore be considered the cream of this crop.

3   Data

To analyse the effect of the intervention by Betfair, we collected data on 11,307 races run in the U.K and Ireland between January 1st 2011 and December 26th 2011.[2] This sample divides almost equally into the pre-treatment period (up to and including June 19th 2011) and the post-treatment period (from June 20th 2011). Table 1 provides a full breakdown of the races. A select group of races are run with only maiden horses (those that are yet to win a race) and novice horses (those that may have a win under their belt, but not this season). Races are also broken down into weekday and weekend races. These four sub-samples are used in our later analysis.

We sourced two types of odds on offer. Those relating to the bookmakers (defined as the starting price, a summary measure of odds at the time the race began) were obtained from Betwise, a betting information company. The betting exchange odds were extracted via the Betfair Application Programmers Interface, and consist of the odds at the time the race began. The exchange odds were 'back' odds, and so are directly comparable to the bookmaker's odds, except that the counterparty was another bettor on the exchange rather than a bookmaker.

Our first task is to establish that the premium charge removed at least some of the most informed from the market. Perhaps the affected bettors simply registered new accounts to evade the charge? Perhaps they had already vacated the market after the first premium charge in 2008?

---

[2]Data on 5,340 further races that took place between 27th December 2011 and 17th June 2012 is used later in this section to replicate our initial findings.

Betfair would be unlikely to implement a charge (with all its operational and public-relations costs) if evasion was simple - and indeed would be unlikely to launch a second charge in 2011 if the first had cleared out all potentially eligible bettors - but, nonetheless, we require statistical evidence that the charge was effective. One way to check if informed bettors did indeed leave the market is to examine the change in odds after the treatment. The market microstructure literature (e.g. Glosten and Milgrom (1985) for financial markets, and Shin (1991, 1992, 1993) for betting markets) would suggest that as the proportion of informed traders declines, *ceteris paribus*, the odds will increase. The idea is that those providing liquidity had previously lost out to informed traders (as they couldn't be distinguished from uninformed traders), so had recouped those losses by keeping the odds low (i.e. keeping the price high) for all traders. This is akin to keeping a wide bid-ask spread, *à la* Glosten and Milgrom (1985) and Copeland and Galai (1983), in standard financial markets. If the informed were to subsequently leave the market, those providing liquidity could then increase the odds and still break even, as they are no longer losing out to the most informed.

Table 2 provides summary statistics on both the bookmaker and betting exchange odds. The betting exchange odds display a greater range, and are consistently higher than those offered by bookmakers. This observation is consistent with Smith *et al.* (2006) who found that returns were higher on Betfair than with bookmakers. More importantly, after the treatment there is a relatively large increase in the average odds quoted on the exchange (from 49.05 to 55.47) yet a smaller increase in the bookmakers' average odds (quoted on the same events) from 20.16 to 20.43. In Table 3 we examine this proposition more formally. We regress the quoted odds (for each horse to win) on an indicator variable equalling 1 if the odds were quoted on Betfair, an indicator variable equalling 1 if the odds were quoted after the treatment, and an interaction between these two variables. The interaction term is crucial as this isolates the effect of the treatment. All regressions in this table involve clustering of standard errors for each race. We find that the treatment did indeed increase the odds on Betfair, which appears to confirm that the adverse selection problem faced by liquidity providers was mitigated by the premium charge, and that at least some informed traders did leave the market. Although the charge only affected approximately 0.1% of the betting population, the resultant increase in the odds is pronounced (an increase of approximately 6 judging by the coefficient associated with the interaction term). Judging by the insignificance of the post-treatment indicator, there was no corresponding increase in bookmaker odds after the charge. In Regression 2 we include control variables - related to the

number of runners and the type and location of the race - and find the result be robust.

Having provided evidence that the charge did remove at least some of the most informed, we turn to the main focus of our analysis: the informational effect. We analyse the forecasting capacity of both the bookmaker and betting exchange odds before and after the treatment on 20th June 2011. We do this by converting the quoted odds into implied probabilities, where $implied win probability = \frac{1}{Odds+1}$, and regressing an indicator variable equalling 1 if the horse won the race, and 0 otherwise, on our calculated implied win probabilities. The R^2 reflects the extent to which variation in our calculated implied probabilities can explain variation in the winner of the race. If explanatory power were absolute (i.e. $R^2 = 1$), the eventual winner would have an implied win probability of 1, with all other horses having an implied win probability of 0.

Table 4 displays the results for our full sample. Examining the change in the (McFadden) Pseudo R^2, we can see that the forecasting capacity of Betfair, the betting exchange, declined by 0.7% after the treatment, in comparison to a decline of 0.5% in the forecasting capacity of bookmaker prices over the same period. This is a relatively small effect, given that a number of highly informed bettors were likely lost after the intervention. We also considered the relative change in forecasting capacity for sub-samples related to location (Britain or Ireland), type of race (e.g. maiden, novice), whether the race was on a weekday or weekend, and whether the race was a handicap race (where superior horses carry heavier weights to level the contest) or not. The following descriptions relate to those sub-samples where the results are discernibly different from the full sample.

Tables 5 and 6 divide races into weekday and weekend races respectively. It could be argued that the most profitable bettors carry out their bets for a career, and therefore their loss would be most keenly felt in weekday races when the majority of the working population is unable to bet. Consistent with this, we find a 0.9% decline in the forecasting capacity of Betfair for this sub-sample. On the other hand, when weekend races are considered, the exchange's forecasting capacity increases by 0.4% compared to a 0.5% decline in the bookmaker control market over the same period. It could therefore be argued that the deterrence of informed traders actually improved informational efficiency in certain cases, as other traders had greater incentives to participate (due to the reduction in adverse selection).

A similar pattern is found in maiden and novice races (see Tables 7 and 8). There is arguably greater potential for bettors to acquire information on novice races, as here the horse may already

have a winning record. In line with this thinking, the forecasting capacity of Betfair improves more than that of the bookmaker control market after Betfair's new premium charge (though the sample size is relatively small here, involving only 408 races). In maiden races, where horses have no winning record - and it is perhaps a little trickier to obtain profitable information - the forecasting capacity of the betting exchange declines relative to the bookmaker control market.

The weekend and novice race results tally with the model of Fishman and Hagerty (1992), where the prohibition of (profitable) insider trading incentivises other traders - who would otherwise have been deterred by the adverse selection problem - to acquire information and reveal their new information to the market. However, as we shall see later in this section, there is no statistical significance to these sub-sample results.

One concern with our methodology is that the Pseudo R^2 is often a noisy indicator of forecasting capacity. We have attempted to mitigate this by measuring the forecasting capacity for the full duration of the pre-treatment and post-treatment periods. This should average-out the majority of the noise. However, it would be instructive to see if we can construct a confidence interval for the Pseudo R^2, so as to have an idea of the actual level of noise in this measure. To this end, we ran the same logit regressions as before, but this time ran them separately for each day in the sample. We were then left with a daily Pseudo R^2. We then calculated a Pseudo R^2 confidence interval for both the bookmaker and the betting exchange, and for both the pre-treatment and post-treatment periods. The 95% confidence interval is quite narrow in all four cases, suggesting that the measure is less noisy than we assumed. For the betting exchange it changes from 0.144987 - 0.1619524 in the pre-treatment period to 0.1382133 - 0.1550493 in the post-treatment period. This slight fall is accompanied by a similar fall in the bookmaker market from 0.1346953 - 0.152152 in the pre-treatment period to 0.1302684 - 0.1476646 in the post-treatment period, and mirrors our earlier results.

Now that we have these daily Pseudo R^2s, we can use them to examine the statistical significance of any change in forecasting capacity. In Table 9 we regress the daily Pseudo R^2 on an indicator variable equalling 1 if the odds were quoted on Betfair, an indicator variable equalling 1 if the day is in the post-treatment period (from June 20th 2011), and an interaction between the two aforementioned terms. As in Table 3, the crucial term is the interaction term, as this captures the effect of the intervention (controlling for confounds using our control market) on the forecasting capacity of the exchange. We find that the intervention had no significant impact on

Betfair's forecasting capacity. This tallies with our results in Table 4, where the relative change was very slight. We also repeat this analysis for the sub-samples discussed earlier, and find that any relative increase (decrease) that we observed in the forecasting capacity of those sub-samples is not statistically significant. These regressions are also displayed in Table 9. To sum up at this stage, although the intervention forced a number of informed bettors off the exchange, the net effect on the forecasting capacity of the exchange was almost nil.

Up until this point, however, our sole measure of the forecasting capacity of the two markets is the change in Pseudo R^2. One alternative way to assess the forecasting capacity of these markets is to calculate the proportion of winners (and losers) that are correctly classified by the logit models. This proportion could then be compared for pre-treatment and post-treatment periods, and across control and treatment markets. The typical threshold for classifying a winner is a win probability of greater than 0.5 inferred from the logit model. However, the resultant statistic from this calculation is not necessarily informative in our setting as the data is heavily skewed. In fact, a model that predicted every horse to lose would be correct approximately 90% of the time as there are, on average, approximately 10 horses in each race.

A solution to this problem is to use a Receiver Operating Characteristic Curve (or ROC curve), as used in Franck *et al.* (2010) to compare the forecasting capacity of a variety of betting markets. This displays the proportion of winners and losers that are correctly classified for the full range of thresholds from 0 to 1. Figure 1 displays this curve for pre-treatment and post-treatment periods for the exchange and the bookmakers. Setting an appropriate threshold is a trade-off between sensitivity (on the vertical axis) which is the proportion of winners correctly classified, and 1 minus specificity (on the horizontal axis) which is the proportion of losers correctly classified. The area under the ROC curve is then calculated to give a measure of the explanatory power of the model. An area of 0.5 indicates no forecasting capacity, while an area of 1 indicates perfect forecasting (akin to an R^2 of 1).

In Table 10 we describe the area under the ROC curve for our full sample and the sub-samples described earlier. For the full sample, the forecasting capacity of the exchange actually fell by less than the control market after the treatment. This supports our earlier assertion that there was no detrimental effect from the tax on the most informed. The sub-sample analysis reveals that, under this new measure, the forecasting capacity of the exchange declined slightly (relative to the control market) for maiden races, where information acquisition is arguably more problematic,

but increased slightly (again, relative to the control market) for novice and weekend races (where information acquisition is perhaps less costly). The only difference in these results from our earlier measure is that there is no relative decline in forecasting capacity for weekday races, where information acquisition (and participation) is arguably more difficult for ordinary bettors.

One other issue with our work so far is that we have only considered the effect of this treatment immediately after its imposition. The 0.1% of traders affected by the tax may take time to move their operations elsewhere. This could explain why there is no perceptible drop-off in the forecasting capacity of the exchange. To examine whether this is the case, we ran our analysis again, but this time used a later treatment period running from 27th December 2011 to 17th June 2012. This includes 5340 new races. The results of this estimation are displayed in Table 11. We observed a 0.03% decline in the forecasting capacity of the exchange relative to a 0.06% increase in the control market. The area under the ROC curve fell to 79.03% (a decline of 0.12%) for the exchange, compared to a larger decline to 79% (a fall of 0.15%) for the control bookmaker market. Even with a lag, there is little evidence of a substantial deleterious effect on the forecasting capacity of the exchange as a result of the intervention.

One final issue is that of inter-market arbitrage. It is conceivable that after the informed bettors left the market, arbitrageurs corrected any subsequent mispricings between Betfair and its competitors. This would ensure that price discovery (see, for example, Hasbrouck (1995)) was not taking place on Betfair, but arbitrage retained the informational efficiency of the market. There is sure to be an element of this, but two factors will likely limit its scope. The first is that there are 'limits' to the arbitrage of price differences between the exchanges. A bettor must pay commission on their net winnings on each exchange, and therefore the price discrepancy must be relatively substantial before it is profitable for an arbitrageur to step in. This first point is exacerbated by the high levels of price volatility that precede the beginning of each race. Even if the arbitrage trade is attractive enough to overcome the commission, the arbitrageur has to be reasonably certain that they can execute the trade (across the two exchanges) before the prices move. It is therefore likely that arbitrage opportunities will often persist across the exchanges.

4   Conclusion

There has long been a focus in the financial and betting market literatures on the actions of 'insiders' or informed traders, and the consequences of their actions (see Vaughan Williams (2005) for a survey). While it is thought that the presence of informed traders leads to lower liquidity (Copeland and Galai (1983) and Glosten and Milgrom (1985)), these traders are also thought to play a crucial role in the price discovery process, as their trades drive the prices of assets towards their fundamental value (Glosten and Milgrom (1985)).

We use a natural experiment to focus, primarily, on the second proposition. In June 2011, Betfair, a betting exchange in the U.K., enacted a rule change which drastically affected the terms of trade for their most profitable, and implicitly most informed, bettors. These bettors would henceforth be subject to a tax of up to 60% on their future profits, leading some to, it appears, vacate this market. We examine the effect that Betfair's intervention had on the informational efficiency of the market - using the bookmaker market as our control group - and thereby attempt to shed light on the way that information is distributed among speculative market participants.

The forecasting capacity of the market (our measure of the market's incorporation of information into the price) declined insignificantly after the rule change. This result is robust to the choice of forecasting measure, the choice of forecasting interval, the choice of post-treatment period, and the choice of race type. This suggests that information was relatively dispersed amongst market participants, and that more than a few traders typically play a role in the price discovery process in speculative markets.

References

- Barber, B., M., Odean, T., (2000). Trading Is Hazardous to Your Wealth: The Common Stock Investment Performance of Individual Investors. *Journal of Finance*, 55, 773-805.

- Barber, B., M., Lee, Y., T., Liu, Y., J., Odean, T., (2014). The Cross-Section of Speculator Skill: Evidence from Day Trading. *Journal of Financial Markets*, 18, 1-24.

- Copeland, T., Galai, D., (1983). Information Effects and the Bid-Ask Spread. *Journal of Finance*, 38, 287-305.

- Coval, J., D., Hirshleifer, D., A., Shumway, T., (2005). Can Individual Investors Beat the Market? School of Finance, Harvard University, Working Paper No. 04-025.

- Engelberg, J., Sasseville, C., Williams, J., (2012). Market Madness? The Case of Mad Money. *Management Science*, 58, 351-364.

- Fishman, M., J., Hagerty, K., M., (1992). Insider Trading and the Efficiency of Stock Prices. *Rand Journal of Economics*, 23, 106-122.

- Franck, E., Verbeek, E., Nüesch, S., (2010). Prediction Accuracy of Different Market Structures - Bookmakers Versus a Betting Exchange. *International Journal of Forecasting*, 26, 448-459.

- 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.

- Hasbrouck, J., (1995). One Security, Many Markets: Determining the Contributions to Price Discovery. *Journal of Finance*, 50, 1175-1199.

- Kaniel, R., Liu, S., Saar, G., Titman, S., (2012). Individual Investor Trading and Return Patterns Around Earning Announcements. *Journal of Finance*, 67, 639-680.

- Kelley, E., K., Tetlock, P., C., (2013). How Wise are Crowds? Insights from Retail Orders and Stock Returns. *Journal of Finance*, 68, 1229-1265.

- 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.

- Shin, H., S., (1991). Optimal Betting Odds Against Insider Traders. *The Economic Journal*, 101,1179-1185.

- Shin, H., S., (1992). Prices of State Contingent Claims with Insider Traders, and the Favourite-Longshot Bias. *The Economic Journal*, 102, 426-435.

- 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.

- Tetlock, P., C., (2011). All the News that's Fit to Reprint. Do Investors React to Stale Information? *Review of Financial Studies*, 24, 1481-1512.

- 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? *Economic Journal*, 107, 150-158.

Tables and Figures

Table 1: Races	Total	Pre-treatment	Post-treatment
All	11307	5268	6039
GB	9329	4426	4903
Ireland	1978	842	1136
Novice	876	468	408
Maiden	1966	841	1125
Weekday	9330	4348	4982
Weekend	1977	920	1057
Handicap	6057	2867	3190
Non-Handicap	5250	2401	2849

A summary of races in our sample, running from **1/1/11** to **26/12/11**. The post-treatment period begins on **20/6/11**. Novice races are between horses that are yet to win a race in the current season, and Maiden races are between horses that are yet to win a race in any season. Handicap races involve superior horses carrying heavier weights to level the field.

Table 2: Summary Statistics

Bookmaker Odds (SP)	Mean	Median	Max.	Min.	Std.Dev	Skewness	Kurtosis
All (N=112053)	20.31	12	501	0.05	24.89	3.52	22.33
Pre-treatment (N=51196)	20.16	12	301	0.05	23.84	3.11	16.99
Post-treatment (N=60857)	20.43	12	501	0.06	25.74	3.78	25.41
Betting Exchange (Back odds)	Mean	Median	Max.	Min.	Std.Dev	Skewness	Kurtosis
All (N=112053)	52.53	14	999	0.04	130.01	5.02	31.69
Pre-treatment (N=51196)	49.05	14	999	0.05	113.91	5.12	34.89
Post-treatment (N=60857)	55.47	14	999	0.04	142.08	4.84	28.6

Summary statistics on the betting odds available with the bookmakers (in the top panel) and the betting exchange, Betfair (in the bottom panel). The sample is separated into pre-treatment and post-treatment, where post-treatment covers the period following the increase in the Betfair premium charge on 20/6/11.

Table 3: Odds Comparison		
Dependent Variable: Odds	1	2
Intercept	20.16642***	-9.214451***
	(0.1682699)	(1.375911)
Betfair Indicator	29.01326***	28.99934***
	(0.7621907)	(0.7621845)
Post-treatment	0.2696275	0.2984712
	(0.2355087)	(0.4124709)
Betfair Indicator*Post-treatment	6.087455***	6.094111***
	(1.156777)	(1.156877)
No. of Runners		3.106104***
		(0.0918811)
Weekend Indicator		-5.674177***
		(0.6916829)
Maiden Indicator		27.91316***
		(1.192188)
Novice Indicator		36.4795***
		(1.891067)
GB Indicator		1.425194***
		(0.7912846)
Handicap Indicator		-24.66265***
		(0.703709)
R^2	0.0294	0.1034
No. of Observations	224106	224106

The odds (for each horse to win) regressed on an indicator variable equalling 1 if the odds were quoted on Betfair, an indicator variable equalling 1 if the event took place in the post-treatment period (from June 20th 2011), and an interaction between the two aforementioned terms. This interaction term is the key variable as it captures the effect of the treatment on the odds. Control variables related to each race are included in Regression 2. Standard errors (clustered for each race (N=11,307)) are in parentheses and *** indicates significance at the 0.1% level.

Table 4: Forecasting Races

Dependent Variable: Race Win Indicator

	Pre-treatment	Pre-treatment	Control	Post-treatment
	1	2	3	4
Intercept	-3.211537***	-3.189781***	-3.306476***	-3.240611***
	(0.026831)	(0.023103)	(0.027022)	(0.021786)
Bookie Implied Win Probability	8.014975***		8.7736***	
	(0.184549)		(0.193906)	
Betfair Implied Win Probability		7.223553***		7.64606***
		(0.116081)		(0.114923)
No. of Observations where Dep. Var.=1	5279	5279	6043	6043
No. of Observations	51196	51196	60857	60857
Pseudo R^2	0.13304	0.1453	0.127671	0.138301
Change in Pseudo R^2			-0.005369	-0.006999

Binary Logit regressions of an indicator variable equalling 1 if the horse won the race (and 0 otherwise), on the implied win probability, as inferred from the odds provided by the bookmakers (our control market) and the betting exchange, Betfair (our treatment market). The pre-treatment phase runs until 19/6/11, with the post-treatment phase beginning after the introduction of the higher premium charge on Betfair on 20/6/11. The change in explanatory power (i.e. change in Pseudo R^2) after this treatment is compared for our two groups. Huber/White standard errors are in parentheses and *** indicates significance at the 0.1% level. This table contains the results for the full sample.

Table 5: Forecasting (Weekday) Races

Dependent Variable: Race Win Indicator

	Pre-treatment	Pre-treatment	Control	Post-treatment
	1	2	3	4
Intercept	-3.194149***	-3.192126***	-3.286285***	-3.226957***
	(0.029273)	(0.025667)	(0.029713)	(0.024075)
Bookie Implied Win Probability	7.903851***		8.648384***	
	(0.199670)		(0.211439)	
Betfair Implied Win Probability		7.255941***		7.564705***
		(0.128619)		(0.125723)
No. of Observations where Dep. Var.=1	4357	4357	4986	4986
No. of Observations	41759	41759	49052	49052
Pseudo R^2	0.133531	0.148337	0.128014	0.138981
Change in Pseudo R^2			-0.005517	-0.009356

Binary Logit regressions of an indicator variable equalling 1 if the horse won the race (and 0 otherwise), on the implied win probability, as inferred from the odds provided by the bookmakers (our control market) and the betting exchange, Betfair (our treatment market). The pre-treatment phase runs until 19/6/11, with the post-treatment phase beginning after the introduction of the higher premium charge on Betfair on 20/6/11. The change in explanatory power (i.e. change in Pseudo R^2) after this treatment is compared for our two groups. Huber/White standard errors are in parentheses and *** indicates significance at the 0.1% level. This table contains the results for weekday races.

Table 6: Forecasting (Weekend) Races

Dependent Variable: Race Win Indicator

	Pre-treatment	Pre-treatment	Control	Post-treatment
	1	2	3	4
Intercept	-3.299466***	-3.178394***	-3.403022***	-3.305776***
	(0.066720)	(0.053010)	(0.064565)	(0.051132)
Bookie Implied Win Probability	8.614916***		9.409816***	
	(0.473947)		(0.480120)	
Betfair Implied Win Probability		7.062632***		8.063745***
		(0.269353)		(0.282038)
No. of Observations where Dep. Var.=1	922	922	1057	1057
No. of Observations	9437	9437	11355	11355
Pseudo R^2	0.130791	0.130833	0.125885	0.134821
Change in Pseudo R^2			-0.004906	0.003988

Binary Logit regressions of an indicator variable equalling 1 if the horse won the race (and 0 otherwise), on the implied win probability, as inferred from the odds provided by the bookmakers (our control market) and the betting exchange, Betfair (our treatment market). The pre-treatment phase runs until 19/6/11, with the post-treatment phase beginning after the introduction of the higher premium charge on Betfair on 20/6/11. The change in explanatory power (i.e. change in Pseudo R^2) after this treatment is compared for our two groups. Huber/White standard errors are in parentheses and *** indicates significance at the 0.1% level. This table contains the results for weekend races.

Table 7: Forecasting (Maiden) Races

Dependent Variable: Race Win Indicator

	Pre-treatment	Pre-treatment	Control	Post-treatment
	1	2	3	4
Intercept	-3.380132***	-3.431336***	-3.408412***	-3.439078***
	(0.058699)	(0.057160)	(0.051291)	(0.048918)
Bookie Implied Win Probability	8.01634***		8.325698***	
	(0.374278)		(0.332094)	
Betfair Implied Win Probability		7.616683***		7.782944***
		(0.254198)		(0.218610)
No. of Observations where Dep. Var.=1	842	842	1125	1125
No. of Observations	8494	8494	12158	12158
Pseudo R^2	0.20214	0.224292	0.196205	0.215972
Change in Pseudo R^2			-0.005935	-0.00832

Binary Logit regressions of an indicator variable equalling 1 if the horse won the race (and 0 otherwise), on the implied win probability, as inferred from the odds provided by the bookmakers (our control market) and the betting exchange, Betfair (our treatment market). The pre-treatment phase runs until 19/6/11, with the post-treatment phase beginning after the introduction of the higher premium charge on Betfair on 20/6/11. The change in explanatory power (i.e. change in Pseudo R^2) after this treatment is compared for our two groups. Huber/White standard errors are in parentheses and *** indicates significance at the 0.1% level. This table contains the results for maiden races, where the horses are yet to win in any season.

Table 8: Forecasting (Novice) Races

Dependent Variable: Race Win Indicator

	Pre-treatment	Pre-treatment	Control	Post-treatment
	1	2	3	4
Intercept	-3.04344***	-3.174716***	-3.145921***	-3.294542***
	(0.073552)	(0.079710)	(0.083989)	(0.089440)
Bookie Implied Win Probability	6.463498***		7.188626***	
	(0.337773)		(0.431213)	
Betfair Implied Win Probability		6.491124***		7.133047***
		(0.283659)		(0.333636)
No. of Observations where Dep. Var.=1	468	468	408	408
No. of Observations	3698	3698	3239	3239
Pseudo R^2	0.207001	0.228529	0.214113	0.23991
Change in Pseudo R^2			0.007112	0.011381

Binary Logit regressions of an indicator variable equalling 1 if the horse won the race (and 0 otherwise), on the implied win probability, as inferred from the odds provided by the bookmakers (our control market) and the betting exchange, Betfair (our treatment market). The pre-treatment phase runs until 19/6/11, with the post-treatment phase beginning after the introduction of the higher premium charge on Betfair on 20/6/11. The change in explanatory power (i.e. change in Pseudo R^2) after this treatment is compared for our two groups. Huber/White standard errors are in parentheses and *** indicates significance at the 0.1% level. This table contains the results for novice races, where the horses are yet to win in the current season.

Table 9: Race Day Pseudo R^2

Dependent Variable: Race Day Pseudo R^2	1	2	3	4	5
	All	Weekday	Weekend	Maiden	Novice
Intercept	0.1434237***	0.1446636***	0.1332261***	0.251042***	0.2359975***
	(0.0044164)	(0.0049281)	(0.0085329)	(0.0161568)	(0.0187999)
Betfair Indicator	0.0100461	0.0115447.	0.0009826	0.0051804	0.0062771
	(0.0061585)	(0.0068068)	(0.0125863)	(0.0222574)	(0.0261418)
Post-treatment	-0.0044572	-0.004574	-0.0015136	-0.0254872	-0.0174192
	(0.0062403)	(0.0069194)	(0.0131155)	(0.0199607)	(0.0262519)
Betfair Indicator*Post-treatment	-0.0023812	-0.0036918	0.0054674	0.0080373	0.0225743
	(0.0086931)	(0.0095949)	(0.0186465)	(0.0277063)	(0.037187)
No. of Observations	662	574	88	621	501
R^2	0.0087	0.0102	0.0029	0.0049	0.0028

The daily Pseudo R^2 (calculated by running the earlier logit regression for each day) regressed on an indicator variable equalling 1 if the odds were quoted on Betfair, an indicator variable equalling 1 if the day is in the post-treatment period (from June 20th 2011), and an interaction between the two aforementioned terms. This interaction term is the key variable as it captures the effect of the treatment on the relative forecasting capacity of the exchange. Regression 1 includes all days in the sample, while Regressions 2, 3, 4 and 5 include weekdays, weekends, days with at least one maiden race, and days with at least one novice race, respectively. White standard errors are in parentheses and *** and . indicates significance at the 0.1% level and 10% level respectively.

Table 10: Logit Model Classification Success		
All	Betfair	Bookie
Pre-treatment	79.15%	79.15%
Post-treatment	78.81%	78.74%
Change	-0.34%	-0.41%
Weekday	Betfair	Bookie
Pre-treatment	79.34%	79.32%
Post-treatment	78.80%	78.74%
Change	-0.54%	-0.58%
Weekend	Betfair	Bookie
Pre-treatment	78.24%	78.33%
Post-treatment	78.84%	78.73%
Change	0.60%	0.40%
Maiden	Betfair	Bookie
Pre-treatment	85.61%	85.66%
Post-treatment	85.17%	85.28%
Change	-0.44%	-0.38%
Novice	Betfair	Bookie
Pre-treatment	85.20%	85.24%
Post-treatment	86.13%	86.08%
Change	0.93%	0.84%

The area under the Receiver Operating Characteristic (ROC) Curve for pre-treatment (1/1/11 - 19/6/11) and post-treatment (20/6/11 - 26/12/11) periods. An area under the ROC curve of 0.5 implies no forecasting capacity while an area of 1 implies perfect forecasting. This is an alternative measure of the explanatory power of the logit model presented in Tables 4-8.

Table 11: Forecasting Races (Treatment Period 2)

Dependent Variable: Race Win Indicator

	Pre-treatment	Pre-treatment	Control	Post-treatment
	1	2	3	4
Intercept	-3.211537***	-3.189781***	-3.24709***	-3.200129***
	(0.026831)	(0.023103)	(0.027449)	(0.023162)
Bookie Implied Win Probability	8.014975***		8.34653***	
	(0.184549)		(0.187210)	
Betfair Implied Win Probability		7.223553***		7.349198***
		(0.116081)		(0.114572)
No. of Observations where Dep. Var.=1	5279	5279	5340	5340
No. of Observations	51196	51196	51324	51324
Pseudo R^2	0.13304	0.1453	0.133721	0.144954
Change in Pseudo R^2			0.000681	-0.000346

Binary Logit regressions of an indicator variable equalling 1 if the horse won the race (and 0 otherwise), on the implied win probability, as inferred from the odds provided by the bookmakers (our control market) and the betting exchange, Betfair (our treatment market). The pre-treatment phase runs from 1/1/11 until 19/6/11, with the post-treatment phase beginning on 27/12/11 and ending on 17/6/12. The change in explanatory power (i.e. change in Pseudo R^2) after this treatment is compared for our two groups. Huber/White standard errors are in parentheses and *** indicates significance at the 0.1% level.
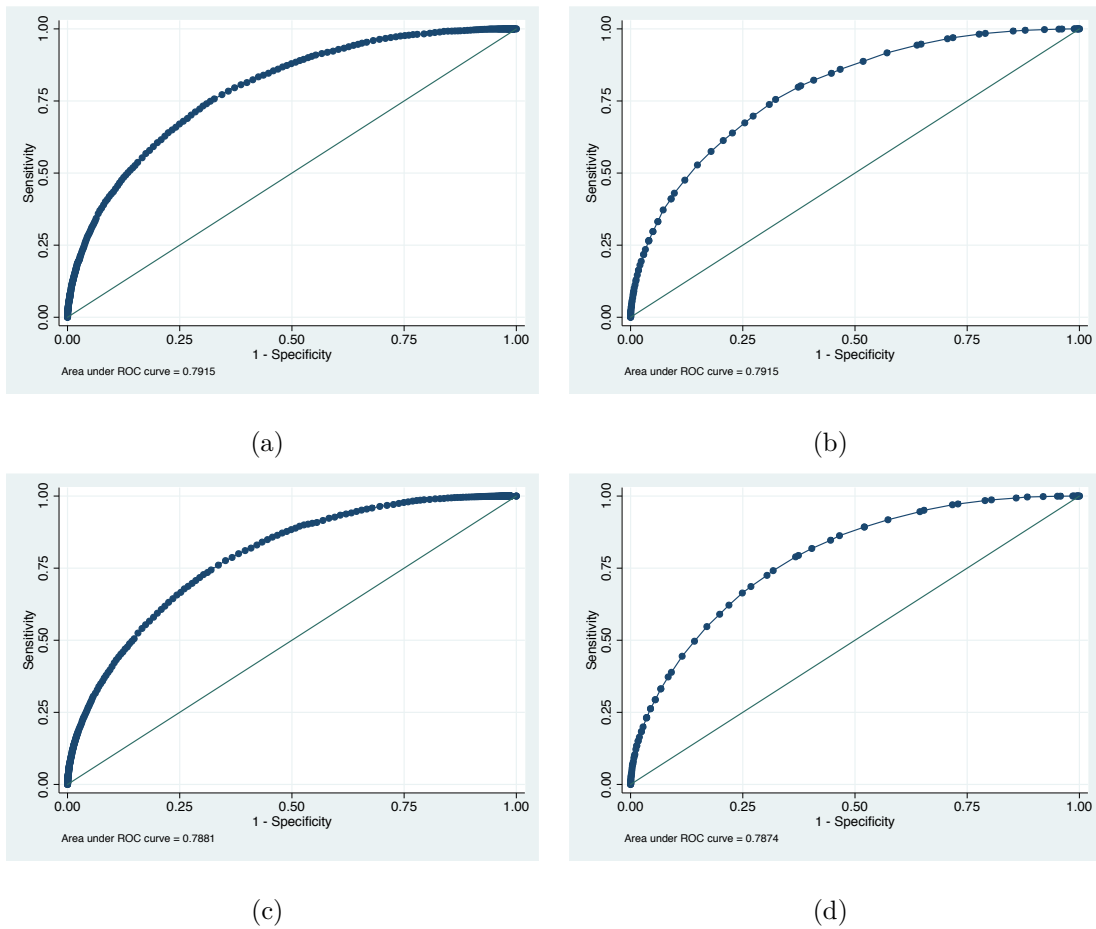
Figure 1: The Receiver Operating Characteristic (ROC) curves which measure the proportion of winners (and losers) correctly classified (for a variety of thresholds) by the logit model described in Table 4. An area under the ROC curve of 0.5 implies no forecasting capacity while an area of 1 implies perfect forecasting. Panels (a) and (b) relate to pre-treatment (1/1/11 - 19/6/11) Betfair and Bookie respectively, while panels (c) and (d) relate to post-treatment (20/6/11 - 26/12/11) Betfair and Bookie respectively.